A HIERARCHICAL DECISION SUPPORT SYSTEM FOR WORKFORCE PLANNING IN MEDICAL EQUIPMENT MAINTENANCE SERVICES

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

ÇİĞDEM CİHANGİR

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
INDUSTRIAL ENGINEERING

DECEMBER 2010

,

Approval of the thesis:

A HIERARCHICAL DECISION SUPPORT SYSTEM FOR WORKFORCE PLANNING IN MEDICAL EQUIPMENT MAINTENANCE SERVICES

submitted by **ÇİĞDEM CİHANGİR** in partial fulfillment of the requirement for the degree of **Master of Science in Industrial Engineering Department, Middle East Technical University** by,

Prof. Dr. Canan Özgen
Dean, Graduate School of **Natural and Applied Sciences** ________________

Prof. Dr. Sinan Kayalıgil ________________
Head of Department, **Industrial Engineering**

Assist. Prof. Dr. Z. Pelin Bayındır ________________
Supervisor, **Industrial Engineering Dept., METU**

Assist. Prof. Dr. Tarkan Tan
Co-Supervisor, **Industrial Engineering Dept.,** ________________
Technical University of Eindhoven, Netherlands

Examining Committee Members

Assist. Prof. Dr. Sedef Meral
Industrial Engineering Dept., METU ________________

Assist. Prof. Dr. Z. Pelin Bayındır ________________
Industrial Engineering Dept., METU

Assist. Prof. Dr. Sinan Gürel
Industrial Engineering Dept., METU ________________

Prof. Dr. Ömer Kırca
Industrial Engineering Dept., METU ________________

Assist. Prof. Dr. Ferda Can Çetinkaya
Industrial Engineering Dept., Çankaya University ________________

Date: **17.12.2010**

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name : Çiğdem CİHANGİR

Signature :

ABSTRACT

A HIERARCHICAL DECISION SUPPORT SYSTEM FOR WORKFORCE PLANNING IN MEDICAL EQUIPMENT MAINTENANCE SERVICES

Cihangir, Çiğdem

M.S., Department of Industrial Engineering

Supervisor : Asst. Prof. Dr. Z. Pelin Bayındır

Co-Supervisor : Asst. Prof. Dr. Tarkan Tan

December 2010, 74 pages

In this thesis, we propose a hierarchical level decision support system for workforce planning in medical equipment maintenance services.

In strategic level, customer clusters and the total number of field engineers is determined via a mixed integer programming and simulation. In MIP, we aim to find the minimum number of field engineers. Afterwards, we analyze service measures such as response time via simulation.

In tactical level, quarterly training program for the field engineers is determined via mixed integer programming and the results are interpreted in terms of service level via simulation.

Keywords: Decision Support System, Hierarchical Workforce Planning, Simulation, Mixed Integer Programming

ÖZ

MEDİKAL EKİPMAN BAKIM VE ONARIM HİZMETLERİ İŞGÜCÜ PLANLAMASI İÇİN HİYERARŞİK YAPIDA KARAR DESTEK SİSTEMİ

Cihangir, Çiğdem

Yüksek Lisans, Endüstri Mühendisliği Bölümü

Tez Yöneticisi : Yrd. Doç. Dr. Z.Pelin Bayındır

Ortak Tez Yöneticisi : Yrd. Doç. Dr. Tarkan Tan

Aralık 2010, 74 Sayfa

Bu tez çalışmasında, medikal ekipman bakım hizmetlerinde insan gücü planlaması için iki hiyerarşik düzeyde karar destek sistemi temeli öneriyoruz.

Karar destek sistemi stratejik ve taktik seviye olmak üzere iki seviyede kurgulandı. Stratejik düzeyde, müşteri kümelerini tespit etmek ve saha mühendisleri sayısını bulmak için karışık tamsayılı programlama kullanıldı. Karışık tamsayılı programlamada analiz edilemeyen servis seviyesiyle ilgili bekleme zamanı gibi servis ölçütleri benzetim yoluyla incelendi.

Taktik düzeyde, saha mühendisleri için üç aylık eğitim planı oluşturmak için karışık tamsayılı programlama ile belirlenen sonuçlar benzetim ile yorumlanıp servis seviyesiyle ilgili ölçütler irdelendi.

Anahtar Kelimeler: Karar Destek Sistemi, Hiyerarşik İşgücü Planlaması, Simülasyon, Karışık Tamsayılı Programlama

ACKNOWLEDGEMENTS

First, I would like to express my deepest gratitude to my supervisors, Assist. Prof. Z. Pelin Bayındır and Assist. Prof Tarkan Tan, for their supervision, patience and understanding. I would like to thank jury members for their valuable contributions on the thesis.

I owe my deepest gratitude to my family; Emel Cihangir, Barbaros Cihangir and Işık Önder for their unconditional support and love.

I would like to thank Eda Göksoy for her wisdom and support throughout my thesis and being a wonderful homemate. I also would like to thank Aycan Öz and Ezgi Çayıroğlu for their support when I was in Eindhoven, for Diclehan Tezcaner and Gülşah Karakaya for their unconditional support and guidance during my thesis. I also would like to thank all my dear friends, especially to Başak Pekdiker, Ezgi Ergül, Alkım Özaygen, Ece Saraoğlu, Nihan Çömden, Özge Narin, Esin Pekpak, Kutay Erbayat for their friendship, support and cheer.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

In recent years, there are a lot of developments and challenges in Service Management area. There is a trend ranging from product-oriented companies to service-oriented businesses. Many services conducted by companies involve going to the customer's location, these services are called field service. Customer satisfaction and loyalty are the key success factors in field service.

In this thesis, we focus on medical equipment maintenance services. The medical equipment service company in concern is responsible for delivering spare parts, upgrades and installations of the systems, technical training of the customers, preventive maintenance and corrective maintenance services. In this study, we restrict our attention to field services of corrective and preventive maintenance and workforce allocation.

Preventive maintenance (PM) is a scheduled activity that has the aim of preventing breakdowns and failures. The main goal of PM is to prevent the failure of equipment before it truly happens. PM is planned once or twice a year for equipments according to the equipment type. For this type of maintenance, field engineers should take an appointment from the customer. Required service level determined by the difference between promised maintenance time and real maintenance time of the PM. PM should be conducted on the day that it is scheduled by a Field Engineer (FE) who can satisfy the skill requirement for that PM task. Corrective Maintenance (CM) is an unplanned activity done upon customer notification when a breakdown of an equipment piece occurs. Service level indicator for CM is response time of the FE with matching the skill requirement.

Knapp and Mahajan (1998) state that manpower planning models for maintenance organizations can be divided into two categories. In supply models, current situation for workforce is analyzed and future changes are predicted and planned. In demand models, workforce required to satisfy the future demand is planned. Also, manpower models can be further examined in two levels. In macro level, aggregate number of people is planned for annual or long-time planning. Micro level decisions are mainly used for personnel matching assignment.

In this study, we aim to build a Decision Support System (DSS) for Field Service Manager of Service Company in concern. We propose a two level hierarchical approach for demand models.

The maintenance service company works in regional customer clusters. Each region has a number of field engineers assigned to the customers in that cluster. There are no base locations for the clusters; field engineers serve from their houses. So that, in strategic level the service company should decide about service clusters and assign a number of field engineers to these clusters. It should be noted that each cluster works independently and the field engineers in a specific service cluster cannot work for any other service clusters.

In strategic level, the Service Company should determine the total number of field engineers and service clysters. So, we build two different Mixed Integer Linear Programming models to model the system. We use these models in order to develop customer clusters and determine minimum workforce for each cluster considering two different objective functions. Afterwards, we use simulation to compare the results. Via simulation, we have an idea of how service level varies by changing the customer clusters and number of field engineers so that Service Company would decide which combination suits best to their motives. We assume that every field engineer is skilled to do all kinds of maintenance activity in this level.

In tactical level, the Service Company should determine the training plan for field engineers quarterly. Initial skill levels and number of field engineers in each cluster is known and in this level. For both PM and CM, there are 7 kinds of product categories and a field engineer can only do the maintenance if he/she is trained for that product category. First, we build a Mixed Integer Linear

Programming Model and use the deterministic data to find feasible training plans. Afterwards, we use simulation to examine the effect of stochastic behavior of the system that we ignore in MIP so that the Service Company would determine which training plan is appropriate.

It should be noted that, we studied medical equipment maintenance services of a service company so that our assumptions are based on the specific characteristics of the company.

In Chapter 2, information about the service company is given; problem definition and solution approach is discussed. After that related literature is examined in three different parts. First, the literature related to the general field service is discussed. And then, the literature that is used in strategic and tactical level is presented.

In Chapter 3, details of strategic level are given. First, two mixed integer programming models are presented. The experimental results of these models are compared. Afterwards, we give information about the simulation model. Lastly, experimental results of the simulation are given and the results are interpreted.

We describe the details of tactical level in Chapter 4. First, mixed integer programming model is presented. After that, the simulation model is given and the result of the MIP Model is analyzed by taking into account the stochastic behavior of the system.

Finally, in Chapter 5, we conclude. Overall results and future research topics are discussed.

CHAPTER 2

PROBLEM DEFINITION AND LITERATURE REVIEW

In this chapter, we first give a brief introduction about the field service operations, information about the company and problem definition. And then we continue with our solution approach. Afterwards, related literature is discussed. We analyze our problem in both strategic and tactical levels; so, we review the previous studies related to our problems described for each hierarchical level.

2.1 Field Service Operations

It is mentioned by Knapp and Majahan (1998) that service organizations have gained importance as profit centers and Field Service Managers have to cope with conflicting objectives such as maintaining a high level of customer service. Therefore, they need tools to analyze the impact of their decisions on customer service and inventory cost. The improvements can be achieved by focusing on improving productivity of service through improved manpower planning. (Papadopoulos, 1996)

2.2 Information about Field Maintenance Service Company

The company of our concern is a part of a multinational company. It has over 300,000 employees world-wide. It is into vast of business segments such as commercial finance, healthcare, energy and industrial.

In organizational structure whole Healthcare Group of the company is divided into two: Devices and Services. Devices section is responsible for research and development and production of the medical equipments.

Service section of the company was founded in 2004 and is responsible for after sales service of the medical equipment devices on the Dutch Market. The company serves approximately 95 hospitals spread over 140 locations and 8 Academic hospitals in the Netherlands.

2.2.1 Organization Structure

The service organization consists of a Service Sale Department, indoor and field services. The service organization is headed by the Service Manager. The office is divided into a repair shop, a call center, and a training coordinator.

*Service Sale Departm*ent is responsible for selling the products and make contracts to the customers. Generally, customers make contracts over different warranty periods with different service level options.

In the Service Company, there is also a repair shop which is responsible for *indoor services*. The products that cannot be repaired in field service are brought to the company and fixed in the repair shop.

The field service consists of field engineers (FE). The customers are divided into four service clusters: North-East, North-West, South-East and South-West. Every FE is assigned to one service cluster and only works for that cluster. Each cluster has a senior FE who coordinates the team in addition to his regular duties. Whole field service is also managed by a Team Leader FE. There is an international base for FEs where they report to and take spare parts they need. Besides that, there is no base location for customer clusters; FEs serve from their houses. So, when a field engineer is hired, the house of the employee is taken into account as he/she will work serve the customers from his/her house.

2.2.2 Maintenance Services

The service organization products and services are related to the products of medical equipments on the Dutch market. The products are grouped into 7 categories; Anesthesia Monitoring, Anesthesia LSS, Bedside Monitoring, Cardiac Monitoring, Cardiac Testing, Infant Care Certification and Telemetry. The service organization provides following products and services:

- Preventive Maintenance
- Corrective maintenance (on field)

- Corrective maintenance (at repair shop)
- Installation
- Calibration
- Technical Training
- Technical Support
- Parts delivery
- Updates / upgrades

These services are offered separately but also in a form of service contracts with or without warranties. In addition, the service company supports the marketing team for demonstration and explanation. All these services are handled by field engineers except corrective maintenance at repair shop which is done by the engineers in the repair shop. In this study, we restrict our attention on field service.

In Figure 1.1, the workloads of field engineers according to the service types for 2004-2008 are shown. It shows that field engineers spend most of their times on preventive and corrective maintenance. Thus, the scope of this project is restricted to preventive maintenance (PM) and corrective maintenance (CM).

For both PM and CM, there are 7 types of product categories: Anesthesia Monitoring, Anesthesia LSS, Bedside Monitoring, Cardiac Monitoring, Cardiac Testing, Infant Care Certification and Telemetry. Every field engineer is trained for one or more product categories. If he/she is not trained, he/she is not capable of doing the maintenance on that product category. So, Field Service Manager has to take in to account the trainings of field engineers while planning.

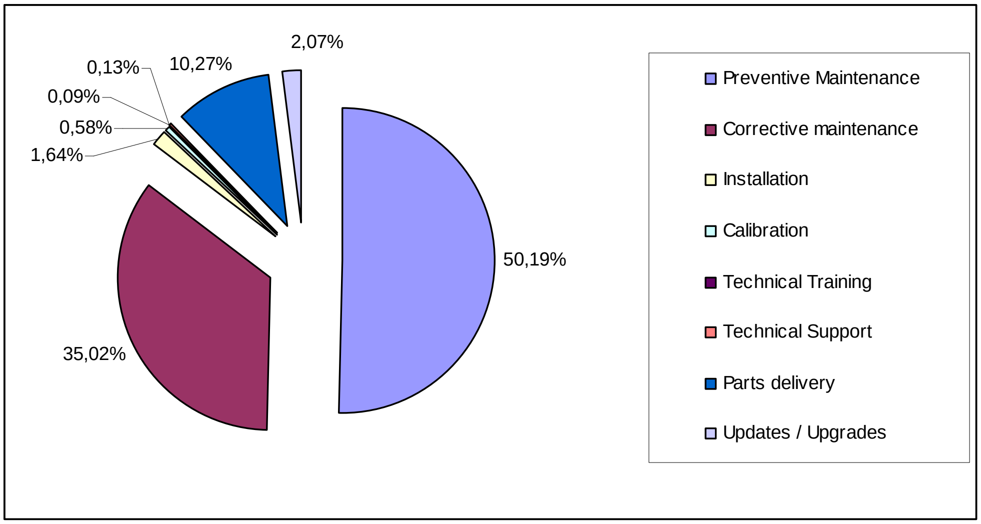

Figure 1.1: Workload Percentage of Field Engineers According to the Services

2.2.2.1 Preventive Maintenance

Preventive maintenance can be both carried out for customers with or without a service contract. Customers with a contract have higher priority. Generally preventive maintenances are once or twice a year and they are yearly planned.

Each team in each customer cluster arranges maintenance tasks in their cluster by themselves. This process is monitored by senior field engineer. Every field engineer makes an appointment according to his own workload and the due months of the preventive maintenances.

Preventive maintenance should be conducted in the days that it is scheduled in order to maintain higher service level.

2.2.2.2 Corrective Maintenance

Generally, corrective maintenance activities are carried out by technical service of hospitals. When technical service of the hospital cannot deal with the situation, they call the Service Company. The maintenance is done by field engineers on field.

When corrective maintenance is needed, firstly, it is tried to be handled by call centers, product specialists via phone. After it is determined that a part should be changed, or customer wants a field engineer to come to the field, FE makes an appointment according to his/her workload and current situation (distance, equipment type etc). It should be noted that, corrective maintenance can be a preemptive activity. It means that if a FE is busy with a PM, he/she can have a break with the PM and go for maintenance regarding to the CM criticality. CM criticality is based on the patient safety; according to priority and criticality of the machine. CMs can be critical or non-critical.

For corrective maintenance, the most important service level indicator is response time of the field engineers.

2.3 Problem Definition

Initial problem statement of the Service Company is “Deliver concrete recommendations how to improve efficiency without losing customer satisfaction”. Starting with this initial statement, our main aim is “To build a Decision Support System in order to recommend improvements to improve efficiency of field service without losing customers satisfaction”. We analyze the problem in two hierarchical levels and design DSS for each level individually.

In strategic level, the service company has to decide service clusters and the number of field engineers for each cluster. It is a strategic level decision because, when a service cluster is built, the field engineers, who are assigned to a specific service cluster, cannot serve for any other service cluster. The service clusters have no base locations so that it is also important for hiring/firing decisions as the field engineers serve from their houses. It also brings complexity to our problem where we deal with building service clusters without knowing the base locations. So, the Service Company has to work with many different cases and evaluate the possible outcomes to compare and choose one of them.

In tactical level, the problem for the Service Company is to decide the training plan of the field engineers for each service cluster. It was mentioned that every field engineer is trained for one or more product categories. If he/she is not trained, he/she is not capable of doing the maintenance on that product category.

In both decisions there are both stochastic and deterministic demands. Preventive maintenances are determined and scheduled beforehand, while the corrective maintenances occur randomly. Main objective is to reduce the operating costs such as wages and training costs whereas increasing or maintaining the desired or required service level and efficiency of the field service operations.

Efficiency of field service operations should be measured in both customer satisfaction and utilization of the field engineers. It should be noted that with limited resources, firms face a trade-off between service efficiency and customer satisfaction i.e. service level. In the literature, customer satisfaction in corrective maintenance is measured as service level, usually defined as response time. Response time is the time between the call of customer for service and the field engineer reach to the customer. Downtime of the equipment that is the time period during which the equipment stays broken is another service level measuring the customer satisfaction. These two measures are dependent on each other. For preventive maintenance, service level indicator is different from CM. PM should be conducted on the day it is scheduled to provide higher service level. Response time is affected by the workload and travel time of Field Engineer. Repair time is affected by complexity of the maintenance, i.e. maintenance type.

2.4 Problem Scope

In order to handle the complexity of the whole system, our main focus in our DSS will be the about manpower planning in field service. Manpower planning is defined as an attempt to match the supply of people with the maintenance tasks available for them and is concerned with having the right number of workers in the right places at the right time.

2.4.1 Narrow System of Interest

Narrow system of interest is the initial system that will be focused on.

For strategic level;

- Preventive Maintenance procedures
- Corrective Maintenance procedures
- Service levels committed in contracts
- Working hours of the Service Company

For tactical level, the following can also be added to strategic level;

- Training procedures in the organization
- Maintenance and equipments criticality measures.

2.4.2 Objectives

For each hierarchical level these objective is specified.

For strategic level, the objective is to minimize the total number of field engineers in MIP. Via simulation, these results are analyzed in detail to compare the service levels. In strategic level, skill levels are not taken into account.

For tactical level, feasible training plans are found by minimizing total relevant costs. The cost values are training costs or opportunity costs such as not serving a maintenance task on time. Afterwards, these training plans are analyzed via simulation to compare the service level observed for each training plan.

2.4.3 Decision Variables and Parameters

Decision variables and parameters for strategic and tactical level can be summarized in Table 2.1 below.

In strategic level, according to the customer locations, repair times and yearly capacity and yearly average demand, service clusters and the number of field engineers for each cluster is found.

In tactical level, according to costs, expected CM for each product type and PM schedules, training plan for field engineers is determined.

Table 2.1: Decision variables and parameters for Strategic and Tactical Level

	Strategic Level	*Tactical Level*
Decision Variables	- Number of field engineers - Service clusters - Allocation of field engineers to the service clusters	- Training plan for field engineers
Parameters	- Yearly capacity of field engineers - Yearly average demand for CM and PM - Locations of customers - Repair times for PM and CM	- Expected CM for each product type - PM schedule for each product type - Cost of serving a maintenance late - Cost of training

2.4.4 Performance Measures

Performance measures are listed below:

- Response time of Field Engineer: Response time is the time period between customer call and field engineer reach to the field area. This is a performance level for corrective maintenance tasks.
- Service level: There are many service level indicators. Most commonly response time, total time in the system are used for corrective maintenance.
- Utilization of Field Engineers
- Travel time of Field Engineers: Field engineers spend their time traveling to the customers and field service operations. It is preferred that field engineers spend less time for traveling to the customer.
- Total relevant costs: Costs can be real costs like wages or opportunity costs like serving a maintenance job late.

2.5 Solution Approach

Main goal of the thesis is to build a DSS to help Service Company make decisions in both strategic and tactical level while sustaining patient safety and maintaining the best coupling for maximum service level and minimum cost.

Proposed project approach is adapting the models in the literature considering unique characteristics of the service company. In this project, demand models will be used for building DSS. In demand models, according to the demand for maintenance tasks, current or future manpower needs are determined (Knapp and Mahajan,1998). In this thesis, the demand models are studied in two hierarchical levels; strategic level, and tactical level planning. Figure 2.1 shows the summary of the solution approach.

Strategic level planning includes the decisions about optimal manpower planning strategies. Via strategic level planning, actual organizational resource and future resource need is highlighted. The decisions are generally decided for a longer period of time; yearly. As the field engineers are divided into the customer clusters, the Service Company should try to build customer clusters and determine the required manpower satisfying the demand in strategic level. At first, the customers are aggregated in order to reduce the computation time by decreasing number of customers. Then, we use two different mixed integer programming models. First model shows the worst case results, where it is assumed that the customers are served from the furthest location in the service cluster and first model is built in order to find the best case solution, where we assume that the service center is in the middle of the cluster. In MIP, we do not examine with service level measures such as response time and waiting time. Thus, we use simulation to imitate the maintenance system and have examined the service levels in detail. Finally, the Service Company can make a decision about building clusters and total number of field engineers in the light of the findings and interpretations of the results.

In tactical level, the objective is to determine the skill levels of the field engineers as well as to have insights about the performance measures such as utilization of field engineers and service level. In our approach, we use mixed integer linear programming and solve the model by using expected values for maintenances to find a feasible training plan for field engineers. Afterwards, the stochastic behavior of the demand is taken into account and we analyze the training plan via simulation. Finally, the Service Company could decide about a training plan for field engineers via interpreting the results.

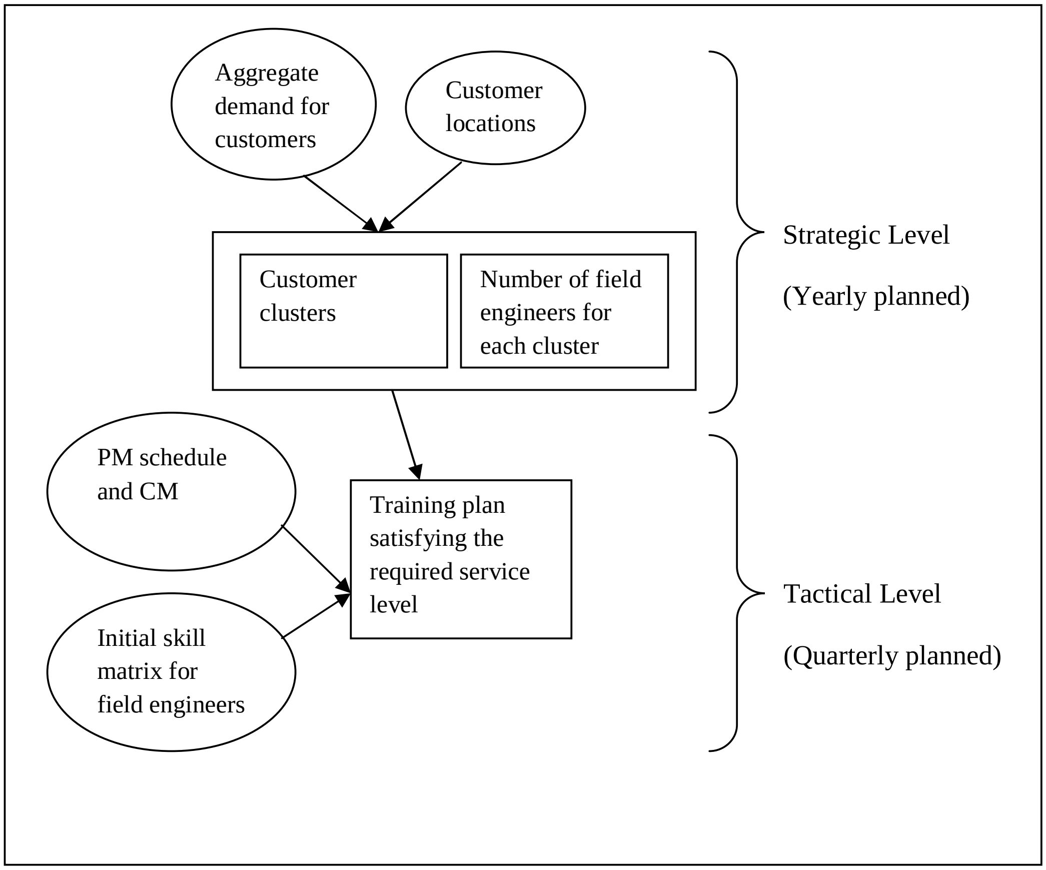

Figure 2.1: Summary of solution approach for strategic and tactical level

2.6 Related Literature

2.6.1 Related Literature about Field Service Operations

A number of articles including Hill (1992) mention that since this field service problem is a complex problem by nature, field service management involves a hierarchy of decisions which include marketing strategy, field service design and operating policies. Lin and Ambler (2005) state that there are three hierarchical decision levels of field technician dispatching in field service design: Strategic planning for national manpower requirements planning, tactical planning for manpower allocation within regions, and operational planning for technician dispatching strategies. In our case, strategic and tactical levels are considered.

Field service planning problems generally includes service area planning and balancing number of field engineers. Service area planning is named as territory planning problem. Territory planning is addressed by many authors for response time performance measure. For instance, Smith (1979) considers the problem of allocating personnel to provide service that involves travel to demand locations distributed over a geographical area. He assumed that a service territory corresponds to all call sources assigned to one server. Within each territory, call locations are assumed to be distributed uniformly with respect to unit area and speed of travel is constant. First-come-first-served dispatching rule is used. He estimates the traveling times assuming the trips can be sequential trips or round trips. He combines these traveling times with the queueing model and illustrates the change between response time and service territory for different kinds of systems. Contrary to Smith (1979), we take into account service clusters more than one field engineer and in our simulation model, we use different types of maintenances (PM and CM) with different criticalities. Travel time is also included in our simulation model in strategic level.

Agnihothri et. al. (2003) remarks that in manpower planning in field service, we have to analyze the trade-off between service response to a customer (customer satisfaction) and service costs. In addition, skill level of the technicians should match the demand. Mostly, simulation-base models are used as tools to analyze that trade-offs and show “what if” analysis. According to Agnihothri et. al. (2002), one of the most important performance measures for field-based services is the down time which is the time between service request and completion of a service. Down time can be divided into response time and on-site time. As Agnihothri et al. (2002) state, the combination of our mixed integer programming models and simulation in strategic level show the trade of between service level and service costs. And it makes these a useful tool for determining the best combination for service level and costs.

Watson et. al. (1998) state that response time is mainly used to have an insight about customer satisfaction. Response time planning involves many sub problems such as requirements planning at the national level, allocation planning at the district level, and operational planning at the team level. We also measured response time via simulation model as well as travel time in our strategic level problem.

Tang et. al. (2008) describe the field service time line as in Figure 2.2. They describe that there are 5 different times in the field service. Queue time is the time between call of the customer and the time when FE begins to travel to the customer. Travel time is the time that FE spends traveling to the customer. Response time is the total of queue and travel times; it is the time between call of customer and the time that FE reaches the customer. On-site repair time is the duration that FE repairs the equipment. Service time is given as the total of the travel and on-site repair time because the capacity of the FE is also spent by traveling.

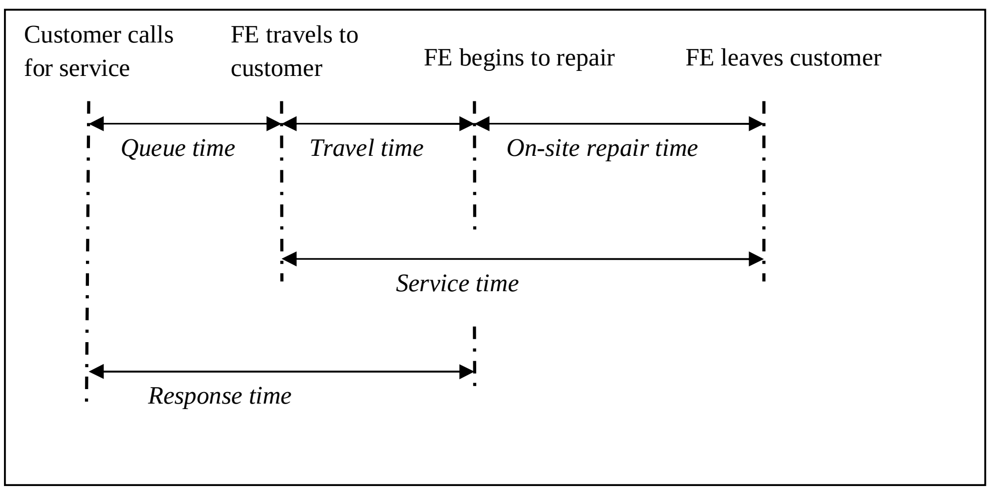

Figure 2.2: Field Service Timeline according to Tang et. al. (2008)

Differing from Tang et. al. (2008), Haugen and Hill (1999) describes service call time line as in Figure 2.3. They show the times more detailed. For example, they consider that there is also a delay time between machine failure and the call of customer.

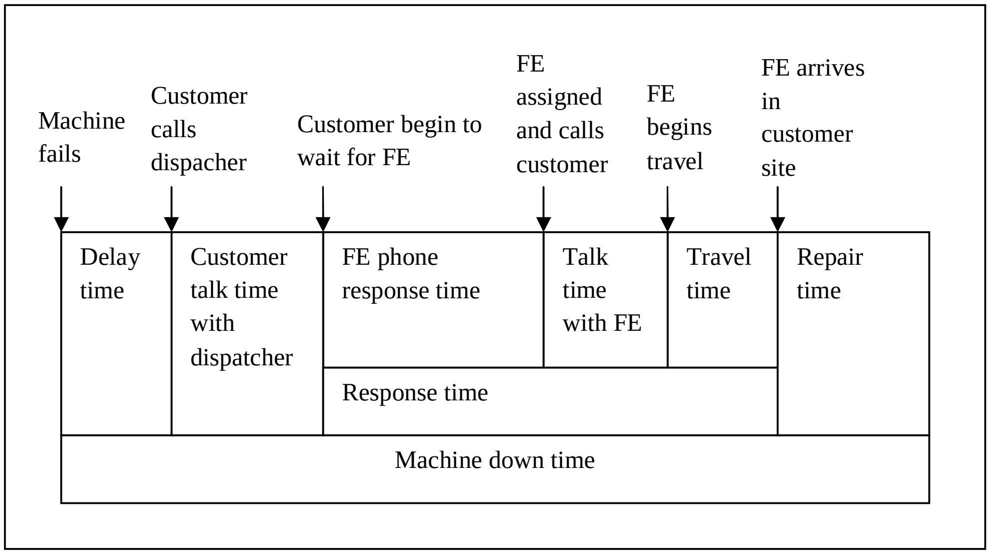

Figure 2.3. Field Service Timeline according to Haugen et. al. (1999)

In their approach, they try to maximize field service quality. They name the problem as traveling technician problem. They address scheduling for different scheduling rules. In their article, they design the system for different dispatching rules such as First-Come-First-Served, Nearest Call, and propose other dispatching rules. They use four base measures of service quality: mean tardiness with respect to service guarantee, mean response time, mean technician phone response time, and mean promise time given at the time the service call is received. To calculate these measures, they use simulation. Hill (1992) proposes several dispatching rules that might be applied for dispatching sequential-trip technicians and compares these rules in a simulation and tries to find out which rule is best in terms of average tardiness. Performance measures used are average queue time, average travel time, average response time, average tardiness, variance of tardiness, and percent of calls that are tardy. In his experiment, different dispatching rules and various problem environments are analyzed. The results give insight about which dispatching rules are more helpful in which situations. In our simulation model, we do not take into account the delay time and we do not use different dispatching rules other than FIFO. However, we use different maintenance types: PM and CM with different criticalities. CMs have a priority over PM where CMs can preempt PMs. We measure response time for each maintenance type individually as they have different service characteristics.

Apte et al. (2007) state that it is a fact that in a non-emergency situation, the customer would prefer to have the service delivered at a time convenient to the customer. For instance, promising a time window of the whole day may seem to be the cost-efficient solution from the viewpoint of a service provider, but for the customer, who requests and expects a shorter time window i.e. exact time to receive the service. We use this approach while considering preventive maintenance. Apte et. al. (2007) also uses simulation in their article. In our simulation in strategic level, we described a preemption number for PMs. This number shows how many times a PM is preempted. Thus, apart from response time, number of preemptions is also a service level for PM.

Hill et al. (1992) and Tang et al. (2008) use state-dependent queueing model to evaluate trade-offs between field service staff level, territory size and response time. They both consider that customer calls arrive according to a Poisson process, and servers (field engineers) are assigned to a given territory. They try to find the minimum number of servers given a target service level percentage of customers calls serviced within time window. Hill et. al. (1992) show that their proposed model gives insights about the trade-off between response times, expected travel times and workforce levels. In our clustering model, we point out trade-off between number of field engineers and number of service clusters as well as total travel time. Like Hill et. al. (1992) and Tang et al. (2008), we examine the trade-off between service cluster size and response time in our simulation model.

2.6.2 Related Literature for Strategic Level Planning

In strategic level, our aim is to build efficient service clusters while minimizing total field engineer number assuming that all engineers are trained flexible. In this problem demand points and average demand values are given for each demand point. In our approach, we first aggregate the demand points in order to reduce the number of customers to solve the clustering problem efficiently. After that, we build service clusters. Finally, we check the service level for each cluster by building a simulation model and interpreting the results on performance measures. We present related literature about this approach in three parts according to the subjects demand aggregation, clustering and simulation approach.

2.6.2.1 Related Literature for Demand Point Aggregation

Francis et al. (2009) state that it is common to aggregate demand points in order to reduce computations in most of the location problems. Aggregation makes the size of the problem more manageable, however every aggregation introduces error. In customer clustering, we are inspired by p-median problems as well as k-means clustering. In p-median location problems the objective is to locate p facilities such as the sum of the distances from each demand point to its nearest facility is minimized. In later Chapters, the customer clustering problem will be discussed more detailed.

According to Gavriliouk (2009) aggregation is basically assignment of several points to a single representative. For this number of distinct aggregate demand points, locations of aggregate demand points and replacement rule to assign to the aggregate demand points must be considered. We also use this approach for aggregation. Francis et. al. (1999) state that the freedom of choice in these three decisions allows numerous aggregation schemes. So, we determined to assign customers to aggregate demand points while trying to balance the workloads of each aggregate demand point.

In aggregation, we also use set covering model as a tool to determine the assignments. The set covering problem is to find a minimum cost set of facilities from among a finite set of candidate facilities so that every demand node is covered by at least one facility. The objective function in set covering problem minimizes the total cost of the facilities that are selected. The simpler version is objective function without cost value. In that version, the model simply minimizes the total number of locations. We also use the simplest version of set covering problem.

According to Hodgson et. al. 2003, aggregation errors for p-median problems can be divided into three sources of error. A and B errors are distance measurement errors. One aggregate demand point represents several demand points. Thus, we need to represent the distance to each facility. Usually, the total weight of the aggregation is assigned to the aggregate demand point, and the resulting weighted distance from the aggregate demand point is used. If this is not exactly the total weighted distance from all demand points in aggregation, then source *A* error occurs. If a facility is located at an aggregated demand point, that aggregate demand, usually an underestimation of the total demand points-facility weighted distance for that aggregation; then source *B* error has occurred. Source *C* error is an allocation error. Allocating an aggregate demand point to the closest facility allocates all of the demand points in that aggregation to that facility. If some of the demand points are allocated to a facility other than their closest, source *C* error has occurred.

However, Francis et. al. (2000) agrees that there seems to be no agreed-upon way to measure aggregation error. In addition, there is no agreed-upon way to compare two different aggregations. That's why; we compare our aggregations by the change in the weighted distances from non-aggregated situation which is simlar to *A* type error.

2.6.2.2 Related Literature for Customer Clustering

For customer clustering, we are inspired by k-means clustering. In k-means clustering problem, we are given a set of data points where each observation is a dimensional real vector, the problem is to partitioning the *n* observations into k sets, so as to minimize the distance from each data point to its nearest center. K-means clustering is NP-hard in general Euclidean space *d* even for 2 clusters (Aloise et. al. 2009 and Dasgupta et. al. 2009). The problem is also NP-hard for a general number of *k* clusters even in the plane (Mahajan et. al. 2009). If *k* and *d* values are fixed, the problem can be solved in time $O\left(n^{dk+1}\log n\right)$, where *n* is the number of entities to be clustered (Inaba et. al. 1994).

K-means algorithm which is proposed by MacQueen, J.B. (1967) is generally used for k-means clustering. Number of clusters k is fixed before starting the algorithm. The algorithm includes 4 steps:

1. K points are placed in the space satisfying they are farthest from each other. These points represent initial group centroids.
2. Each point is assigned to the closest centroid.
3. When all points are assigned, the positions of the K centroids are recalculated.
4. Steps 2 and 3 are repeated until the centroids no longer move.

Although it can be proved that the procedure will always terminate, the k-means algorithm does not necessarily find the most optimal solution; the global objective function minimum.

As mentioned, using well-known clustering algorithm like k-means is common. However, Yaman et. al. (2008) propose an alternative constrained optimization problem formulation. They state that the reason of performing optimization models for clustering instead of existing clustering algorithms is the size of our problem. K-means is generally used for high size problems.

In their article, they present mathematical programming based clustering approach for customer segmentation. Information about the shopping trip of a customer gathered by video records include the total shopping time of that customer, coordinates that she/he visited and time spent at each coordinate. By using this information they try to cluster the shopping behavior of the customers.

The clustering problem is formulated as a mixed-integer programming problem with the objective of minimizing the maximum cluster diameter among all clusters to obtain compact clusters. They assume that desired k clusters are known and the distances are Euclidean. They use three different objective functions: minimizing within cluster pair-wise distances, minimizing maximum diameter of all generated clusters and minimizing sum of all diameters. We also use this approach in our clustering in order to find the near optimal solution to our clustering problem. However, our objective functions are different because of the characteristics of our system.

2.6.2.3 Related Literature for Simulation

Vast of researchers approach field service problem via simulation because of its complex nature. Duffuaa et al. (2001) develop a conceptual simulation model for maintenance systems in plants integrating both preventive and corrective maintenance; i.e. planned and unplanned maintenance. They state that maintenance resources are manpower, materials and spare parts, tools and equipment; and standards and procedures. The conceptual model is developed in order to address the issue of effect of priority levels on scheduling and affect of spare parts policies on down time and optimal manpower requirement.

Duffuaa et al. (2001) state that this conceptual model can be used to develop a stochastic discrete event simulation to study maintenance manpower requirements, spare parts provisioning and impact of different priorities on the costs and specified performance measures. The concern in the article is maintenance systems in plants but it gives insights about maintenance systems and broad view of the maintenance systems. It should be noted that our system is different from the system that Duffua et. al. (2001) described, but they have some similarities. In their conceptual model, Duffuaa et. al. (2001) propose a maintenance service in a factory, field service is not taken into account. So, traveling time is ignored while spare parts inventory is taken into account. In our case, the effect of spare parts is small as field engineers carry the needed materials with them. However, we focus on traveling to customer sites while building our models.

Dear et al. (2000) present a simulation model that shows the evaluation of the resource allocations and dispatching strategies for technicians performing on-site repairs in a copy machine company. In their model, only corrective maintenance is the concern. The basic questions that they try to answer are how many technicians should be employed, where should the technicians should be based, and what dispatching strategy should be used to assign technicians to service calls. In our simulation model, we do not take into account different dispatching rules. However, we describe two different maintenances and the relationship between those maintenances, criticality measures in our model.

Waller (1994) develops a closed queueing network model for field service corrective maintenance. He tries to find out the optimum the number of workers required, the allocation of Field Engineers to customers, and spare part management. In his approach, the managers generally have three primary opportunities to invest capital in the system to improve the service: increase the spare parts inventory, hire more Field Engineers and/or decrease the emergency delivery time. The model shows trade-offs and gives insights and guidance to decide about these issues. In our system, we need to build customer clusters, so that we are not able to find the most appropriate number of field engineer via simulation or queueing model directly and spare parts inventory is out of our scope. However, we manage to observe the relationship between preventive and corrective maintenance.

Papadopoulos (1996) develops a closed queuing network model for field service corrective maintenance incorporating priority classes of customers. In his model, there are various classes of customers, depending on the type of their contract with the company. In our simulation model, we do not distinguish customers; however we distinguish maintenance types as preventive and corrective maintenance and different product categories are considered for tactical level.

Wong (1980) studies preventive maintenance and corrective maintenance in a factory using a simulation technique. He identifies elements of field maintenance system: Technicians, procedure for assigning technicians to perform maintenance, dynamic movements of technicians such as travel and repair times, preventive and corrective maintenance requirements of the facilities. Two types of failures are defined in the model, random facility failure that cannot be prevented by site visitations and failure that shows gradual equipment deterioration. For the latter one, three state Markov model is used: normal state, pending failure and actual failure. According to the model, pending failure state can be spotted and failure can be prevented by technicians. Important performance measures for him are response times, travel times, repair times, number of preventive maintenance performed and cancelled. We also use a similar approach to Wong (1980) in simulation model. However, our system is a field service maintenance system whereas Wong (1980) concern factory maintenance system. Thus, we included traveling time in our simulation model. Apart from that, in strategic level, we also determine the number of field engineers and service clusters so that in addition to observe the results, we also have chance to examine the trade-offs with “what if” analysis via different number of clusters.

2.6.3 Related Literature for Tactical Level

In tactical level, our intention is to build a training plan for the field engineers. First, we use mixed integer linear programming model to find a feasible training plan with given initial skill levels and then we build a simulation model and interpret the results by taking into account stochastic system.

Brusco and Johns (1998) present an integer linear programming model for evaluating cross-training configurations. Their objective is to minimize workforce staffing costs subject to the labor requirements. The model was used to evaluate cross-training structures across different labor requirement patterns for a large paper mill. The results indicate that cross-training structures are preferable. Their findings indicate that partially cross-training employees can result in significant cost savings for a service delivery system. This article gives insights about how important cross-training is.

In their article, Gomar et. al. (2002) develop a linear programming model in order to optimize cross-trained workforce assignment in a construction project. The model suggests how many and what type of worker should be hired and fired and when to switch workers to other activity. Their objective is to minimize weighted sum of total number of workers, hires and fires, and switching. Switching is defined as changing the activity of the worker. It is observed that cross-trained workforce is mostly preferred. The model also gives insights about the assignments of workers to jobs, idle time, switching from job to job and hires and fires.

Srour et. al. (2002) studies workforce training and allocation on construction projects. Their aim is to reduce costs, make best use of the resources, and improve schedule performance of the project. Demand during the project and available workers are given. The model helps in hiring/firing and training decisions. Inputs are available labor, cost of trainings, cost of hiring, labor costs, and estimates about the skill levels. Cross-trained worker is distinguished into two: A worker can posses a primary skill and a secondary skill or know two skills equally. Objective is to minimize labor costs while satisfying project labor demand. The model recommends hiring as many cross-trained workers as possible as expected because cross-trained workers can be assigned to more than one type job on different periods.

Both Gomar et. al. (2002) and Srour et. al. (2002) studied over cross training for construction projects where the decision maker can choose between workers and the works can be fired and new employees can be hired during the project period. In our case, we use an initial workforce and cannot hire and fire field engineers. The Service Company determines training decisions. Srour et. al. (2002) also take into account training of the workers whereas they still have firing and hiring option.

In our simulation model for tactical level, we also used the insights and experiences that we gained in literature for strategic level simulation model. Unfortunately, there are no other simulations that point out the different type of maintenances and different kind of product categories at the same time. There are some simulations which consider more than one type of maintenances. For example, Visser and Howes (2007) develop a discrete event simulation model to investigate a service company which provides planned and breakdown maintenance to hospital and clinics for medical imaging equipment. Their simulation model aims to determine the number of technicians that would optimize profit. Three main output variables are total technician time needed per day, total chargeable time per day, total contract time (non-chargeable) time for the day. In their simulation, the contracted jobs and non-contract jobs are distinguished because contract jobs are assumed to not generating extra revenue to the service company.

CHAPTER 3

STRATEGIC LEVEL DECISIONS

In this Chapter, strategic level decisions are discussed. As it is mentioned in Chapter 2 the field engineers work in regional bases with service clusters. Thus, in this Chapter, we present a decision support system for forming the service clusters and calculating the total number of field engineers. We develop a mixed integer linear programming model for clustering and determining the minimum workforce for each cluster. Different objectives are considered in the model. In Section 1, the problem formulation is presented. Firstly, the customers are aggregated. Afterwards, the clustering model is solved for different objectives. For the strategic level planning, we consider two scenarios. In Scenario 1, it is assumed that the current data for the demand will not change. In Scenario 2, the demand of the customers is increased by 10% which is the expected increase rate for service demand. The increase in demand is applied to the aggregated demand values. The DSS is applied to the Service Company in Netherlands. Finally, solutions according to different objective functions and different scenarios are discussed and the results are given.

3.1 Aggregation of Customers

Service Company totally has 188 customers in Netherlands in 115 municipalities. We consider that the customers in the same city can be represented as a single customer, since the maximum distance within two customers in the same city is 20 km (approximately 0.5 hrs). Therefore, the customers of company are represented by 115 customer locations in our models.

We use the past four years (2004-2008) data to find the workloads of customers. Average number of corrective and preventive maintenances for each of the customer is considered:

$$\text{Expected Service Demand} = \text{Expected demand for PM} + \text{Expected demand for CM}$$

Expected demand for PM (CM) is found by,

$$\begin{pmatrix}\text{Expected demand of} \\ \text{PM (CM)}\end{pmatrix} = \begin{pmatrix}\text{Expected number of} \\ \text{PM (CM)}\end{pmatrix} \cdot \begin{pmatrix}\text{Average demand for a} \\ \text{PM (CM)}\end{pmatrix}$$

Then, we find all the workloads of customers. It is verified that the workloads of customers are normally distributed. Lower confidence interval for %90 is constructed and the customers which have less workload than the lower confidence interval limit are ignored. Finally, we continue with 103 customers and their demands.

When we try to solve the clustering model with the current data, high computational time is incurred. We build a MIP for aggregation and our main objective in aggregation problem is to reduce computational time of the clustering model while building aggregate demand centers with balanced workload. In current situation the workloads of customers are not balanced. In addition, we observe that some of the customers are very close to each other so that they can be aggregated into one customer. With these motivations, in order to reduce the number of customers, we try to aggregate the customers within a determined time distance with each other.

We use a modified version of Covering Problem to model our aggregation problem. In our approach, the customers are assigned to customers that are chosen as the aggregate demand points. At the same time, we try to balance the workloads of aggregate demand points. Afterwards, travel durations within the aggregate demand points are calculated and added to the workloads. After that, we continue the clustering problem with aggregate demand points.

3.1.1 Aggregation Model

Sets

J : Set of customers, where $J = \{1,...,103\}$

Indices

i , j : Indices for customers and $i \in J$ where $J = \{1,...,103\}$

Parameters

p : Number of aggregate demand points

d : Covering distance

$t_{i,j}$: Time distance between customer i and j

$$a_{i,j} = \begin{cases} 1 \text{ if the distance between customer } i \text{ and } j \text{ is smaller than covering distance } (t_{i,j} \leq d) \\ 0 \text{ Otherwise} \end{cases}$$

w_i : Average workload hours of customer i

Decision Variables

$$x_{i,j} = \begin{cases} 1 & \text{if customer } i \text{ is assigned to aggregate demand point } j \\ 0 & \text{Otherwise} \end{cases}$$

$$y_j = \begin{cases} 1 \text{ if customer } j \text{ is chosen as aggregate demand point} \\ 0 \text{ Otherwise} \end{cases}$$

ε : Maximum difference between workload of aggregate demand points and average workload

Objective Function

$$Min \quad \varepsilon \qquad (0)$$

Subject to

$$\sum_j x_{i,j} = 1 \quad \text{for } \forall\, i \qquad (1)$$

$$x_{i,j} \leq a_{i,j} \text{ for } \forall i,\, j \qquad (2)$$

$$\sum_j y_j = p \qquad (3)$$

$$x_{i,j} \leq y_j \text{ for } \forall i, j \qquad (4)$$

$$\sum_i x_{i,j} \cdot w_i \leq \frac{\sum_i w_i}{p} + \varepsilon \quad \text{ for } \forall j \qquad (5)$$

$$\varepsilon \geq 0 \qquad (6)$$

$$x_{i,j}, y_j \in \{0,1\} \text{ for } \forall i, j \qquad (7)$$

In this model, objective function (0) minimizes maximum difference between workload of aggregate demand point and average workload.

Constraint set (1) ensures that all the customers are assigned to an aggregate demand point. Constraint set (2) provides that a customer can only be assigned to an aggregate demand point covering it, i.e. the distance in time between the aggregate demand point and the customer is smaller than given covering distance *d*. The number of aggregate demand points is given as a parameter. So constraint set (3) assures this statement. Constraint set (4) ensures that every aggregate demand point is assigned to itself. Constraint set (5) is used to find ε value. Constraint sets (6) and (7) are non-negativity and binary constraints.

After we find the aggregate demand points and which customers that are assigned to those aggregate demand points, we use the formula below to find the demand of each aggregate demand point regarding the travel times within each aggregate demand point.

$$w_j^{agg} = \sum_i x_{i,j} \cdot (w_i + t_{i,j} \cdot v_i) \quad \text{for } \forall j$$

where

w_j^{agg} : Total demand of aggregate demand point *j* regarding travel times within

$t_{i,j}$: Time distance between customer *i* and aggregate demand point j

v_i : Average number of visits to customer *i*

This problem is solved for different values of *d* and *p.* The solution of this problem will be chosen by the decision maker, i.e. Service Company. Then it will be used as an input for clustering problem by the decision maker considering workload difference between aggregated customers (ε) and covering distance (*d*).

3.1.2 Experimental Results for Aggregation

We solve the model for different combinations of number of aggregation points (*p*) and covering distance (*d*) to obtain the minimum workload difference between aggregated customers (ε).

In our calculations, we consider 103 customers in the Netherlands. As it is mentioned that the workloads of customers in the same city are combined and customers in the same city are assumed to be one customer. In Figure 5, each pin represents a customer in the Netherlands.

In our model, our aim is to build aggregate demand centers with balanced workload. We try to achieve this by minimizing epsilon value which is the maximum difference between workload of aggregate demand points and average workload. Number of aggregate demand points and covering distance in time are the parameters for the model. We solve the model for different parameter values and keep track of the epsilon value to retrieve the best aggregate demand points.

The changes of notation according to different covering distances, 0.2, 0.25, 0.3, 0.4, 0.5 and 0.6 hours are shown in Figures 3.2. As it can be seen in Figure 3.2, when we increase the number of aggregate demand points, we first see rapid decrease and a steady increase in epsilon. The lines indicate the results that the problem is feasible for that covering distance. As expected, feasible region is smaller in smaller covering distance.

Figure 3.1: Customers in Netherlands

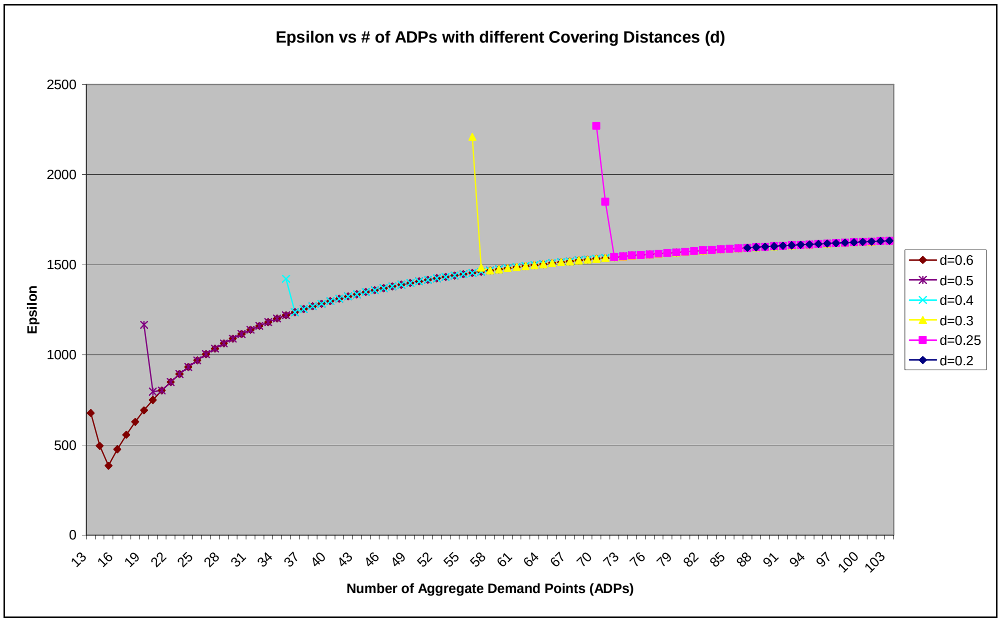

Figure 3.2: Epsilon vs. Number of Aggregate Demand Points (ADPs) with Different Covering Distance

The summary of the results are shown in Table 3.1. Minimum epsilon values are shown in bold. These values are assumed to be best objective function values for that covering distance.

In Figure 3.3, these best combinations are shown in a graph. For the best combinations; it can be seen that, as *d* increases, epsilon and number of aggregate demand points decrease.

Table 3.1: Epsilon Values for Different Number of Aggregate Demand Points (ADP) and Covering Distance *d*

Covering distance in hours (*d*)								
0.2	ADP no (*p*)	86	***87***	88	89	90	91	92
	ε	Infeasible	***1594.365***	1597.229	1600.286	1602.766	1605.443	1608.626
0.25	ADP no (*p*)	69	70	71	***72***	73	74	75
	ε	Infeasible	2270.655	1850.119	***1541.856***	1546.283	1550.875	1554.384
0.3	ADP no (*p*)	55	56	57	***58***	59	60	61
	ε	Infeasible	2210.425	1484.575	***1468.345***	1474.752	1480.947	1486.938
0.4	ADP no (*p*)	34	35	***36***	37	38	39	40
	ε	Infeasible	1421.832	***1237.308***	1253.77	1269.366	1284.161	1298.218
0.5	ADP no (*p*)	18	19	***20***	21	22	23	24
	ε	Infeasible	1166.686	***795.6075***	802.2384	849.7005	893.3546	932.7592
0.6	ADP no (*p*)	12	13	14	***15***	16	17	18
	ε	Infeasible	677.9892	494.9149	***384.5717***	475.9363	556.5521	628.2106

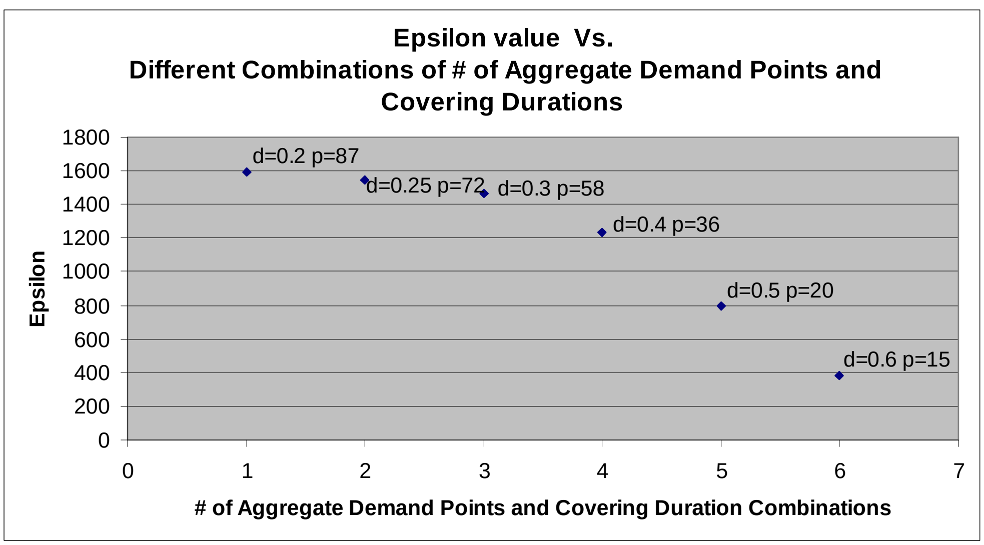

Figure 3.3: Epsilon vs. Different combinations of Number of Aggregate Demand Points and Covering Durations.

It is mentioned by Francis et al. (2000) that there is no agreed-upon way to measure aggregation error. Thus, we calculate the median values for disaggregated data and aggregated data for different *p* and *d* combinations.

First, we calculate the median point for disaggregated data. Then we find all the median points for the best combinations in Figure 3.3. Finally, we calculate the Euclidean distance between the median of best combinations and median of disaggregated data to compare each other. We used these values to have an idea upon the aggregation error. In Figure 3.4, the differences of the medians are shown. As supposed, aggregation error increases when the number of aggregate demand points (*p*) increases.

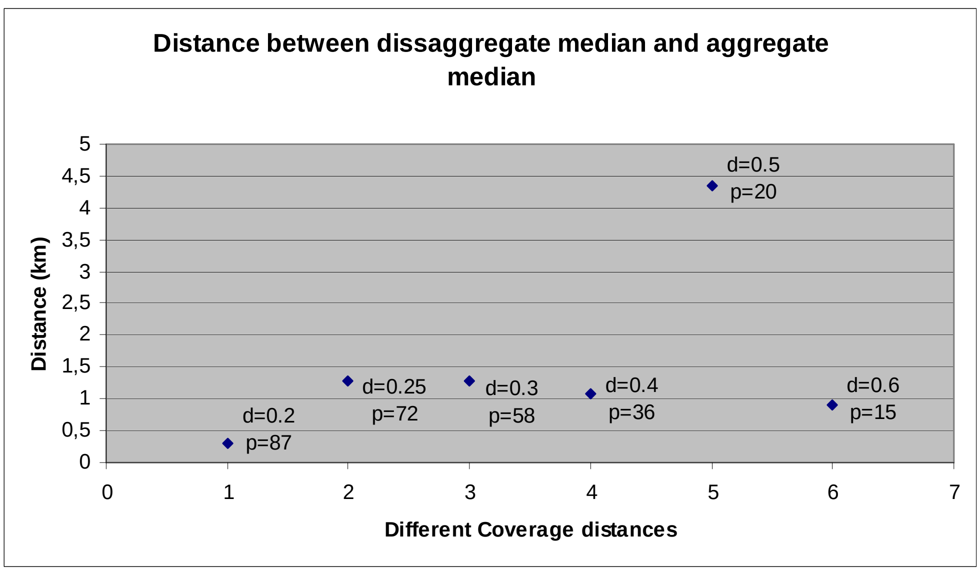

Figure 3.4: Distance between disaggregate median and aggregate median

Figure 3.4 shows that there is no relationship between the median values. All of the values are near to each other except 0.5 hours. In this case, we choose to continue to clustering problem with the pair and $p = 15$ and $d = 0.6$.

In order to get more stable results in clustering problem, we duplicate the aggregate demand points. We divide the workloads by average workload and duplicate some of the aggregate demand points in order to balance the workloads.

Finally, we have 26 customers to continue with clustering problem.

3.2 Clustering Problem

In clustering problem, we try to build customer clusters. The Service Company works in customer clusters and there are no base locations for these clusters; the field engineers serve customers from their houses. The houses of the field engineers are not known before strategic level decisions are made so we use two different models with different assumptions to model.

In the first model, the worst case, we assume that all the customers in a cluster are served from the furthermost customer in that cluster. In the second model, we assume that there is a virtual cluster center in each cluster and all the customers are served from there.

The basic assumptions employed in modeling are as follows:

- All FEs are assumed to be trained in every possible skill.
- There is a fixed cost including the transportation costs for each field engineer. So, total number of field engineers determines the cost. Thus, our objective is *to minimize total number of engineers.* It should be noted that while calculating the total number of field engineers, expected transportation time is considered as well as the expected workload.
- At most 10 clusters can be formed.
- In order to obtain more efficient results, upper bound for the maximum time distance between customers in the cluster is added to the model. The customers that are farther than the specified upper bound are forced to be in different clusters.

- Yearly capacity of a field engineers is assumed to be 2080 hours. ($260 \text{ days} \cdot 8 \text{ hours/day} = 2080 \text{ hours}$)

3.2.1 Model 1

In Model 1, it is assumed that each customer is served from the furthermost customer in the cluster according to its average visit number.

Sets

J': Set of aggregate customers, where $J' = \{1,...,26\}$

K : Set for cluster numbers, where $K = 1,.., p$

Indices

i, j : Indices for customers and $i, j \in J'$ where $J' = \{1,...,26\}$

k : Index for clusters and $k \in K$ where $K = 1,.., p$

Parameters

p : Number of clusters

w_i : Workload of customer *i*

$t_{i,j}$: Time spent traveling from customer *i* to customer *j*

W : Total yearly available working hours of a FE

v_i : Average number of visits for customer *i*

Decision Variables

$$x_{i,k} = \begin{cases} 1 & \text{if customer } i \text{ is assigned to cluster } k \\ 0 & \text{Otherwise} \end{cases}$$

s_k : Number of FE in cluster *k*

$t_{i,k}^{\max}$: Maximum distance from customer *i* to another customer in cluster *k*

Objective Function

$$Min\sum_{k=1}^{p} s_k \qquad (0-1)$$

Subject to

$$\sum_{k=1}^{p} x_{i,k} = 1 \quad \text{for } \forall\, i \qquad (1-1)$$

$$t_{i,k}^{\max} \geq (x_{i,k} + x_{j,k} - 1)\cdot t_{i,j} \quad \text{for } \forall\, i, j, k \qquad (2-1)$$

$$s_k \geq \frac{\sum_i \left(w_i \cdot x_{i,k} + t_{i,k}^{\max} \cdot v_i\right)}{W} \quad \text{for } \forall\, k \qquad (3-1)$$

$$x_{i,k} + x_{j,k} \leq 1 \quad \text{for } \forall\, i, j, k \text{ where } p > 1 \text{ and } \; t_{i,j} \geq 2.5\,\text{hrs} \qquad (4-1)$$

$$s_k, t_{i,k}^{\max} \geq 0 \text{ and } s_k \forall k \;\; \text{integer} \qquad (5-1)$$

$$x_{i,k} \in \{0,1\} \qquad (6-1)$$

Objective function is described in (0-1) as minimizing total number of field engineers. Constraint set (1-1) provides that each customer is assigned to a cluster. Constraint set (2-1) provides that if customer *i* and customer *j* are in the same cluster the distance between them is counted, otherwise the right hand side becomes “0”. The expression on left hand side value is the maximum time spent while going to another customer from customer *i*. Constraint (3-1) calculates the workload for each cluster regarding the worst case situation assumption and provides number of field engineers for each cluster. We assume that. Thus, Constraint (4-1) ensures that the customers which have a time distance greater than or equal to 2.5 hours with each other cannot be in the same cluster. Constraint (5-1) and (6-1) are non-negativity and binary constraints.

3.2.2 Model 2

In this model, one of the customers is chosen to be the cluster center and the customers are assumed to be served from that cluster center according to their expected number of visits.

Sets

J': Set of aggregate customers, where $J' = 1,...,26$

K: Set for cluster numbers, where $K = 1,.., p$

Indices

i, j : Indices for customers and $i, j \in J'$ where $J' = 1,...,26$

k : Index for clusters and $k \in K$ where $K = 1,.., p$

Parameters

p : Number of clusters

w_i : Workload of customer i

$t_{i,j}$: Time spent traveling from customer i to customer j

W : Total yearly available working hours of a FE

v_i : Average number of visits for customer i

Decision Variables

s_k : Number of FEs in cluster k

$$x_{i,k} = \begin{cases} 1 & \text{if customer } i \text{ is assigned to cluster } k \\ 0 & \text{Otherwise} \end{cases}$$

$$c_{i,k} = \begin{cases} 1 & \text{if customer } i \text{ is cluster center of cluster } k \\ 0 & \text{Otherwise} \end{cases}$$

$d_{i,k}$: Time spent traveling from cluster center to customer i if customer i is in cluster k

Objective Function

$$Min \sum_{k} s_k \qquad (0-2)$$

Subject to

$$\sum_{k=1}^{p} x_{i,k} = 1 \quad \text{for } \forall\, i \qquad (1-2)$$

$$\sum_{i} c_{i,k} = 1 \quad \text{for } \forall\, k \qquad (2-2)$$

$$\sum_{k} c_{i,k} \leq 1 \quad \text{for } \forall\, i \qquad (3-2)$$

$$c_{i,k} - x_{i,k} \leq 0 \quad \text{for } \forall\, i,k \qquad (4-2)$$

$$d_{i.k} \geq \sum_{j} (c_{j,k} + x_{i,k} - 1) \cdot v_i \cdot t_{i,j} \quad \text{for } \forall\, i,k \qquad (5-2)$$

$$s_k \geq \frac{\sum_{i} \left(w_i \cdot x_{i,k} + d_{i.k} \right)}{W} \quad \text{for } \forall\, k \qquad (6-2)$$

$$x_{i,k} + x_{j,k} \leq 1 \quad \text{for } \forall\, i, j, k \text{ where } t_{i,j} \geq 2.5 \text{ hrs} \qquad (7-2)$$

$$d_{i.k}, s_k \geq 0 \quad \text{for } \forall\, i,k \text{ and } s_k \forall k \text{ integer} \qquad (8-2)$$

$$x_{i,k}, c_{i.,k} \in \{0,1\} \qquad (9-2)$$

Objective function (0-2), Constraint sets (1-2) and (7-2) are the same for Objective function for Model 1 (0-1), Constraint sets (1-1) and (4-1), respectively. Constraint set (2-2) provides that each cluster has a cluster center. Constraint set (3-2) assures that each customer can be a cluster center at most once. Constraint set (4-2) assures that a customer can be a cluster center if it is assigned to that cluster. Time spent traveling from cluster center to customer *i* if customer i is in cluster k is found by Constraint sets (5-2) and (6-2) calculate the demand of each cluster and determines the field engineer number that should be assigned to that cluster. Constraint sets (8-2) and (9-2) are non-negativity and binary constraints.

The model is solved by CPLEX 10.1. CPLEX is run for at most 18000 seconds (5 hours). The computer that is used for the calculations has Intel Core2 Duo 2.00 GHz, 3 GB RAM. Table 3.2 shows the CPU times and relative gaps for each model and each cluster number under Scenario 1 and 2. In Scenario 1, total workload is assumed to be average workload whereas in Scenario 2 the total workload is decreased by 10%. It should be noted that we run the models for more than two days, and we could not find the optimal solution for 5, 6, 7, 8, 9 and 10 clusters.

It is significant to find the total number of field engineers for each cluster and to show the service clusters however fine tuning of the solutions for service clusters can be defined by the Service Company later. So, we aim to find a near optimal solution where we compare the results.

Table 3.2: CPU Times and Relative Gaps for Scenario 1 and 2

Model No	Number of Clusters	Scenario 1		Scenario 2	
		CPU Time (Seconds)	Relative Gap	CPU Time (Seconds)	Relative Gap
Model 1	1 Cluster	0.03	0%	0	0%
	2 Clusters	0.17	0%	0.39	0%
	3 Clusters	20.92	0%	1.93	0%
	4 Clusters	45.99	0%	246.12	0%
	5 Clusters	18000	12.23	18000	10.12%
	6 Clusters	18000	14.03%	18000	15.66%
	7 Clusters	18000	16.32%	18000	17.56%
	8 Clusters	18000	16.34%	18000	17.13%
	9 Clusters	18000	16.72%	18000	18.07%
	10 Clusters	18000	17.05%	18000	16.32%
Model 2	1 Cluster	0.03	0%	0	0%
	2 Clusters	0.23	0%	0.05	0%
	3 Clusters	0.84	0%	19.73	0%
	4 Clusters	22.13	0%	70.67	0%
	5 Clusters	18000	9.95%	18000	7.35%
	6 Clusters	18000	10.02%	18000	8.45%
	7 Clusters	18000	10.07%	18000	8.97%
	8 Clusters	18000	10.12%	18000	8.76%
	9 Clusters	18000	11.16%	18000	9.08%
	10 Clusters	18000	11.45%	18000	9.10%

Apart from the number of field engineers, utilization of the field engineers and total travel times are other performance measures. For each model and number of cluster, these performance measures are calculated. Average utilizations, travel time percentages, average travel times per cluster are calculated as follows:

$$\text{Average utilization} = \left(\frac{\text{Total Travel Time} + \text{Total Maintenance Workload}}{\text{Number of Field Engineers} \cdot \begin{pmatrix} \text{Yearly Capacity} \\ \text{of a Field Engineer} \end{pmatrix}} \right) \cdot 100$$

$$\text{Travel Time Percentage} = \left(\frac{\text{Total Travel Time}}{\text{Total Travel Time} + \text{Total Maintenance Workload}} \right) \cdot 100$$

$$\text{Average Travel Time per Cluster} = \left(\frac{\text{Total Travel Time}}{\text{Number of Clusters}} \right)$$

The summary of the results for Scenario 1 are shown in Table 3.3. In Figures 3.5, 3.6 and 3.7, the performance measures, total travel time, travel time percentage and total number of field engineers, for Model 1 and 2 under Scenario 1 and 2 are depicted. As expected, total travel time decreases as the number of clusters increases. In addition, average travel time per cluster, and travel time percentage also decrease when number of clusters increases for both of the models. Number of field engineers shows a rapid decrease with the increase in the number of clusters. Afterwards, number of field engineers becomes steady with the increase in the number of clusters. As expected, number of field engineers for Model 1 is greater than the number of field engineers in Model 2. Similarly, the results for Scenario 2 are shown in Table 3.4.

The results show that total travel time decreases when the number of clusters decreases. Numbers of field engineers required are the same when the number of clusters is greater than 3. We can have an insight about the general behavior of the system however we should keep in mind that all the engineers are assumed to be trained in every product category and the stochastic behavior of the system is not taken into account. In addition, the results of the mixed integer models are found with a relative gap from the optimum values. Considering all of these facts, we build a simulation model to compare these clusters more detailed and observe the performance measures such as waiting times of maintenances and average time that a maintenance task spends in system.

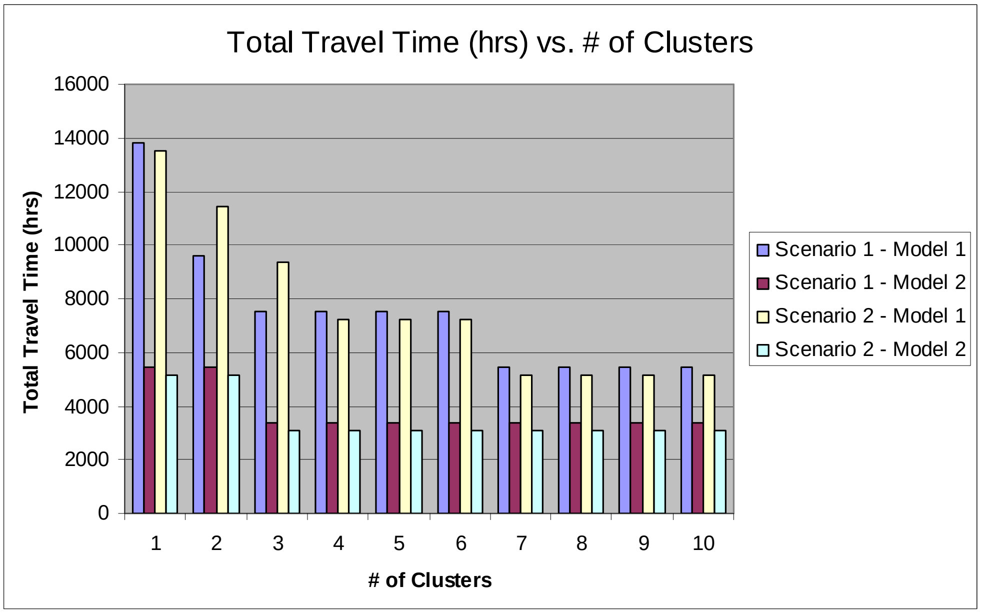

Figure 3.5: Total Travel Time vs Number of Clusters for Each Model and Scenario

Table 3.3: Results for Clustering Models For Scenario 1.

MODEL 1										
Number of Clusters	1	2	3	4	5	6	7	8	9	10
Average Utilization	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
Travel Time Percentage	36.8%	28.9%	24.2%	24.2%	24.2%	24.2%	18.8%	18.8%	18.8%	18.8%
Total Travel Time	13784.39	9624.39	7544.39	7544.39	7544.39	7544.39	5464.39	5464.39	5464.39	5464.39
Average Travel Time per Cluster	13784.39	4812.19	2514.79	1886.09	1508.87	1257.39	780.62	683.04	607.15	546.43
Total Number of FE	18	16	15	15	15	15	14	14	14	14

MODEL 2										
Number of Clusters	1	2	3	4	5	6	7	8	9	10
Average Utilization	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Travel Time Percentage	18.8%	18.8%	12.5%	12.5%	12.5%	12.5%	12.5%	12.5%	12.5%	12.5%
Total Travel Time	5464.392	5464.392	3384.392	3384.392	3384.392	3384.392	3384.392	3384.392	3384.392	3384.392
Average Travel Time per Cluster	5464.392	2732.196	1128.131	846.0981	676.8785	564.0654	483.4846	423.049	376.0436	338.4392
Total Number of FE	14	14	13	13	13	13	13	13	13	13

Table 3.4: Results for Clustering Models For Scenario 2.

MODEL 1										
Number of Clusters	1	2	3	4	5	6	7	8	9	10
Average Utilization	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
Travel Time Percentage	34.2%	30.5%	26.4%	21.8%	21.8%	21.8%	16.6%	16.6%	16.6%	16.6%
Total Travel Time	13498.83	11418.83	9338.83	7258.83	7258.83	7258.83	5178.83	5178.83	5178.83	5178.83
Average Travel Time per Cluster	13498.83	5709.42	3112.94	1814.71	1451.77	1209.81	739.83	647.35	575.43	517.88
Total Number of FE	19	18	17	16	16	16	15	15	15	15

MODEL 2										
Number of Clusters	1	2	3	4	5	6	7	8	9	10
Average Utilization	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Travel Time Percentage	16.6%	16.6%	10.6%	10.6%	10.6%	10.6%	10.6%	10.6%	10.6%	10.6%
Total Travel Time	5178.83	5178.83	3098.83	3098.83	3098.83	3098.83	3098.83	3098.83	3098.83	3098.83
Average Travel Time per Cluster	5178.83	2589.42	1032.94	774.71	619.77	516.47	442.69	387.35	344.31	309.88
Total Number of FE	15	15	14	14	14	14	14	14	14	14

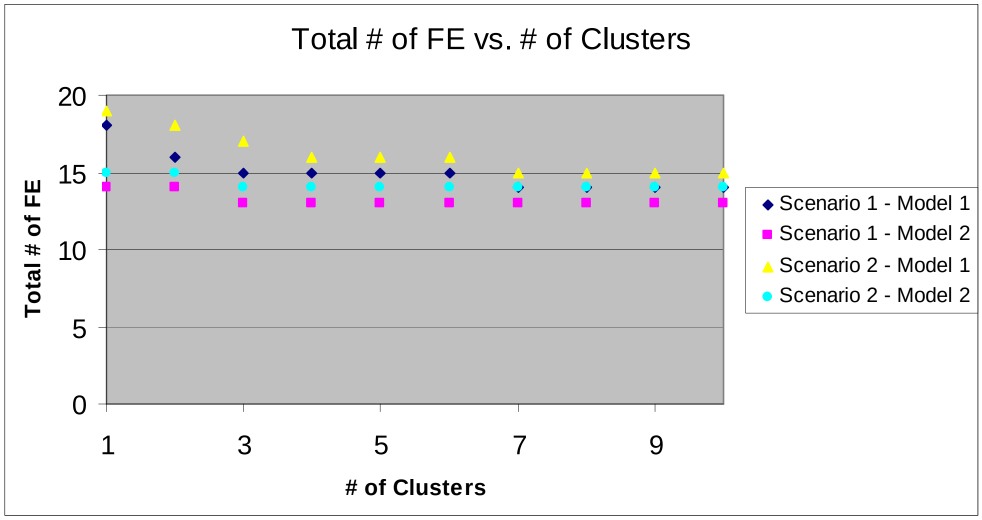

Figure 3.6: Total Number of FE vs Number of Clusters for Each Model and Scenario

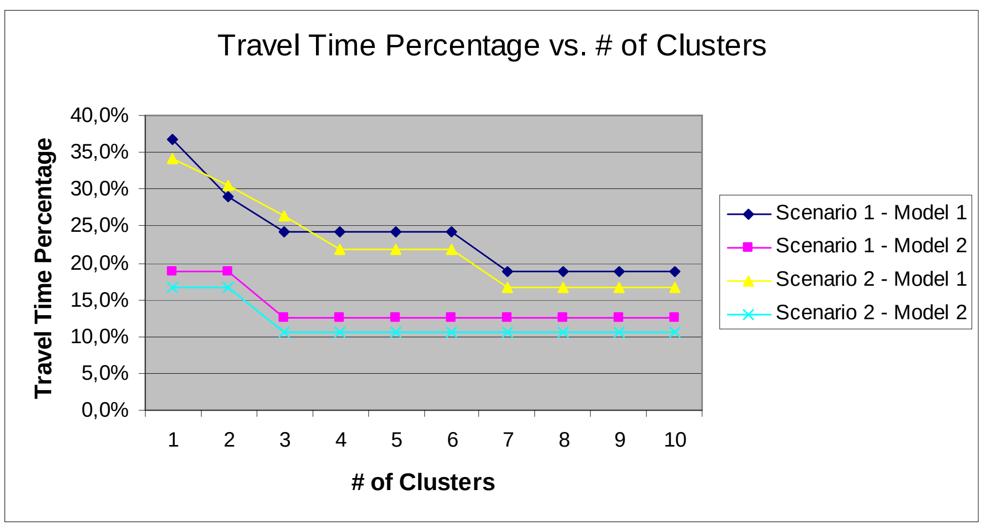

Figure 3.7: Total Time Percentage vs Number of Clusters for Each Model and Scenario

3.3 Simulation Model

We construct a simulation model in ARENA in order to analyze the results more realistically and find out more about the service level; waiting times and response times for preventive and corrective maintenances. In MIP, we find the results under deterministic demand and can only observe total number of field engineers and travel time as performance measures. In real life, the demand for corrective maintenance is stochastic and we need to analyze service level performance measures such as waiting time of customers, response time and total time in system. Thus, simulation is the appropriate tool to analyze the system.

In our simulation model, our focus is preventive maintenances (PM) and corrective maintenances (CM). We design the system so that a CM can preempt PM. Number of preemptions that a PM is preempted is also recorded as well as travel times and repair times for each maintenance type. Utilization of FE, waiting times of maintenances in queues and total time that a maintenance order spends in the system are also recorded. Repair times of CM and PM, arrival of CM and PM, travel times in the clusters are random variables in the model. In MIP, we find the customer clusters; so, according to the number of visits and time distance between customers we find random distributions for travel time.

The flowchart that describes the simulation model can be seen in Figure 3.8.

After a PM or CM is created with exponential distributions with lambda 2.2 and 3.4 respectively, and travel time and repair time is assigned according to the relevant distributions. Model 1 and Model 2 differs in the sense of assumption for travel times. As it was mentioned, Model 1 assumes that each customer is served from the furthest customer for each cluster whereas Model 2 assumes that each customer is served from the cluster center. So, the travel time distributions are different for each model type and each cluster. For each cluster, workloads of the cluster regardless of travel time are known. These workloads ratios are assumed to be probabilities that a PM or CM is in that cluster. So, after the assignment of travel and repair times, the maintenance is assigned to a cluster and sent to the PM or CM queue of that cluster. Every cluster has two queues: PM and CM queue. It should be noted that in PM queue, a maintenance which has a lower arrival time has more priority, i.e. FIFO.

After an activity starts, FE travels to the customer and does the maintenance. So in the model, there is a delay of travel and maintenance after the queues. CMs preempt PMs and how many times that a PM is preempted is kept track. After preemption, PM is sent to PM-queue again. Also, if a PM is preempted, the field engineer has to delay for the travel time, but he can continue the maintenance later. After maintenance activity is finished, the maintenance entity is disposed from the system.

For each cluster, a different model is constructed in order to analyze the dynamics of each cluster independently.

3.3.1 Parameter Estimation

The parameters and how they are derived are listed in following:

- Arrival rate

Arrival rates for CM and PM are derived from 4-years data (2004-2008). For PM and CM, the probability distribution is found to be exponential with lambda 2.25 and 3.40 hrs, respectively.

- Repair time

Repair times for CM and PM are also derived from 4-years data (2004-2008). Repair time distributions for CM and PM are shown in Table 3.5.

Table 3.5: Repair time distributions

Maintenance Type	Repair time distribution
PM	20.5 + 3 * BETA(1.18, 0.921)
CM	3.5 + 2 * BETA(1.06, 0.967)

- Travel time

Travel time distributions are found from MIP solutions. Different distributions are derived for each model and for each cluster.

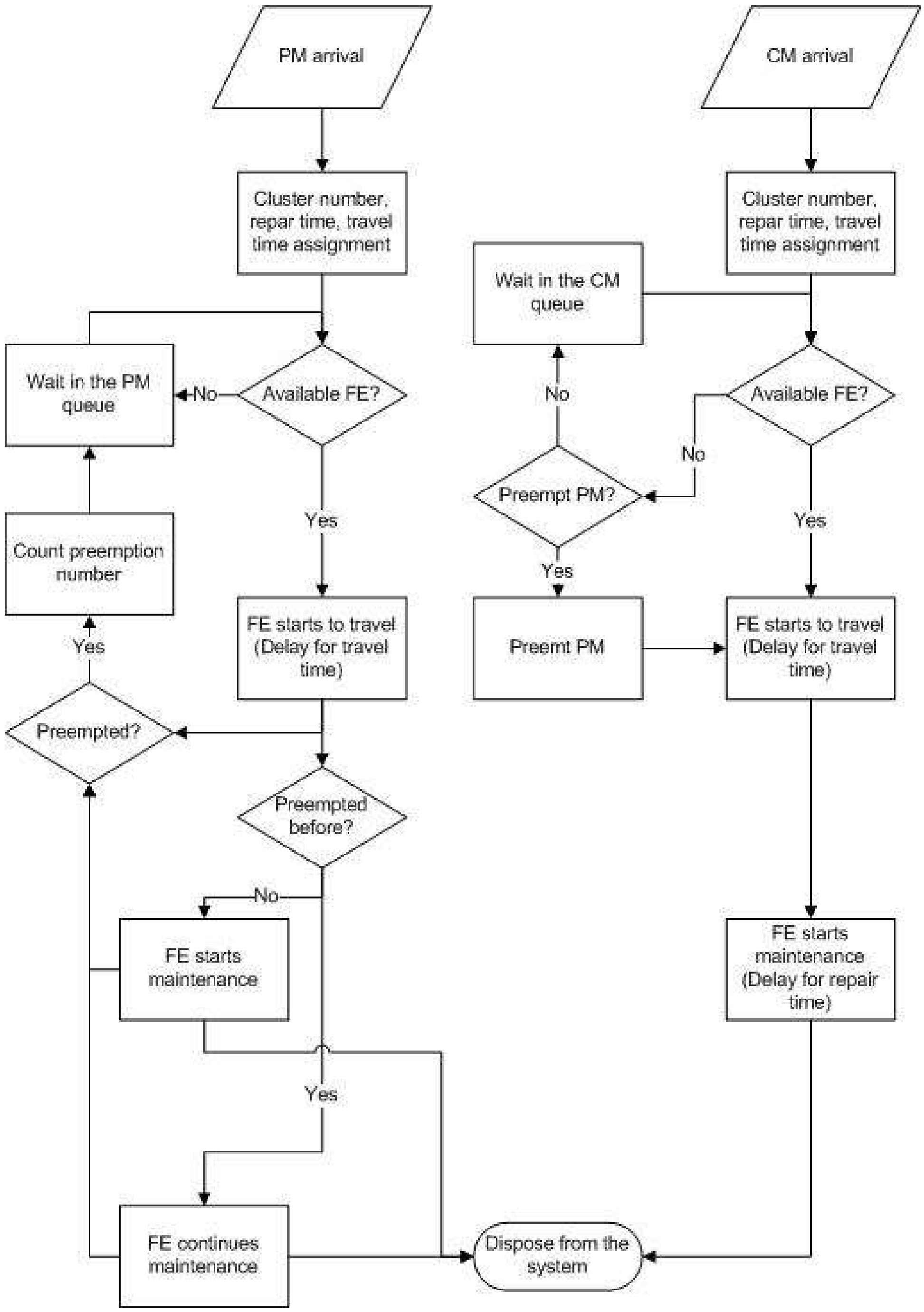

Figure 3.8: Flowchart of Simulation Model for Strategic Level

- Cluster workload percentages

  Cluster workload percentages are also derived from MIP solutions. Different workloads are incurred for each model and for each cluster solution.

3.3.2 Modules in Simulation Model

In this part, each module of the simulation model is discussed in detail. Examples given are for one cluster case.

3.3.2.1 Preventive Maintenance

Preventive maintenance module is shown in Figure 3.9. First, preventive maintenances are created exponentially. It is found that time between arrivals for preventive maintenance is distributed exponentially with lambda value of 2.25 hours. After that the maintenances are counted for verification. After they arrive in the system, travel time, repair time and types are assigned. If there is more than 1 cluster, there is a branch block before assignment which distributes maintenances to different clusters with a probability that is found according to the workloads of clusters. Travel time distribution is unique for each cluster and each model, so that before running the simulation model for each MIP solution, we make sure that right assignment block is active. Repair time distribution differs for preventive and corrective maintenance. Finally, type value is assigned to show that these maintenances are preventive maintenance. Afterwards, the maintenances go to the PM-queue.

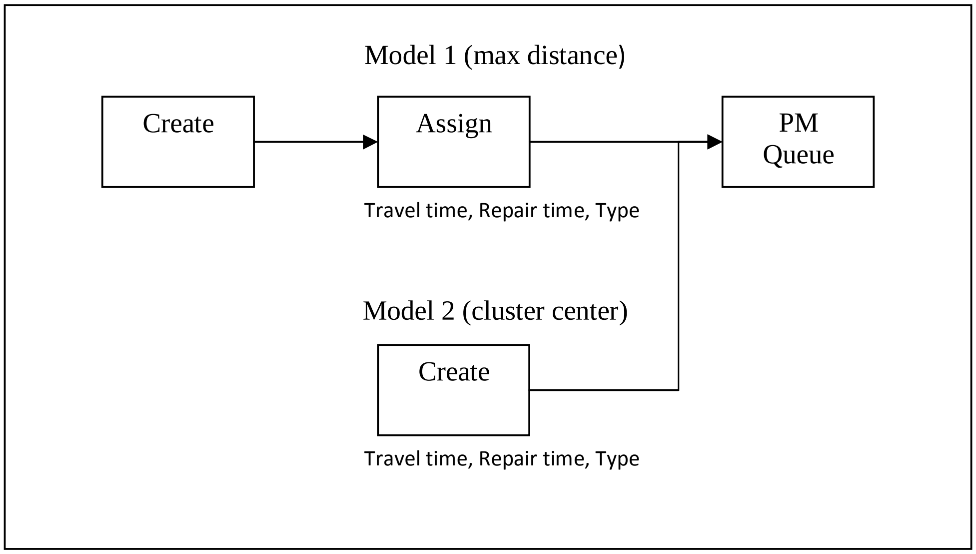

Figure 3.9: Preventive Maintenance Module

3.3.2.2 Corrective Maintenance

Corrective maintenance module is shown in Figure 3.10. The flow is similar to the preventive maintenance. First, corrective maintenances are created exponentially. Time between arrivals for corrective maintenance is distributed exponentially with lambda value of 3.40 hours. Afterwards, travel time, repair time and type are assigned. Repair time is smaller than preventive maintenance. Finally, the maintenances go to the PM-queue.

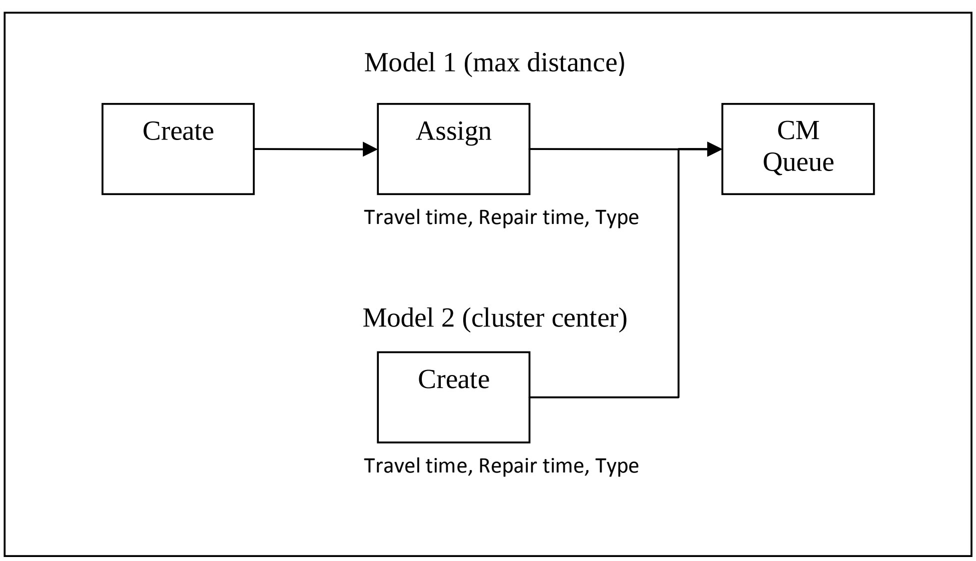

Figure 3.10: Corrective Maintenance Module

3.3.2.3 Preemption Module

Preemption module is shown in Figure 3.11. After queue, PMs are seized with First-In-First-Out rule if there is a field engineer (FE) available. However, when a CM is in the queue, FE should leave the PM activity and do the CM. So, CMs preempt PMs. It should be noted that if a PM is preempted, FE have to travel to the PM location again; whereas, he/she can continue with the repair. When a PM is preempted, how many times that it is preempted is also counted. As it will be further analyzed, in order to satisfy the finite number of tally blocks, preemptions are counted up to “5”. All other values those are bigger than “5” are assumed to be “6”. Afterwards, they are sent to PM-Queue.

When a PM is served, the time between preemption time and served time is also recorded by tally block to give insights about the service level for PMs.

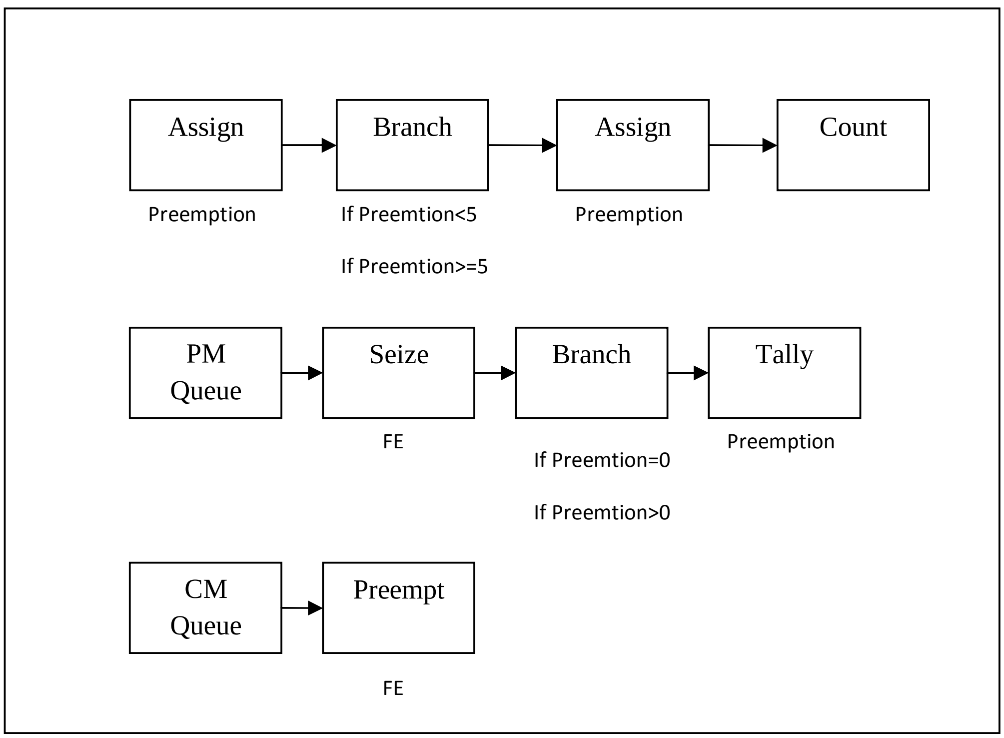

Figure 3.11: Preemption Module

3.3.2.4 Service Module

Service module is shown in Figure 3.12. After a CM preempts a PM or a PM is served, first FE travels to the service location. All field engineers are assumed to know every skill for this hierarchical level. Travel time for PM and CM is recorded. After the travel, he/she starts to repair. Repair times for PM and CM is also recorded. After the service, total time in system for each maintenance type is calculated and the maintenances leave the system.

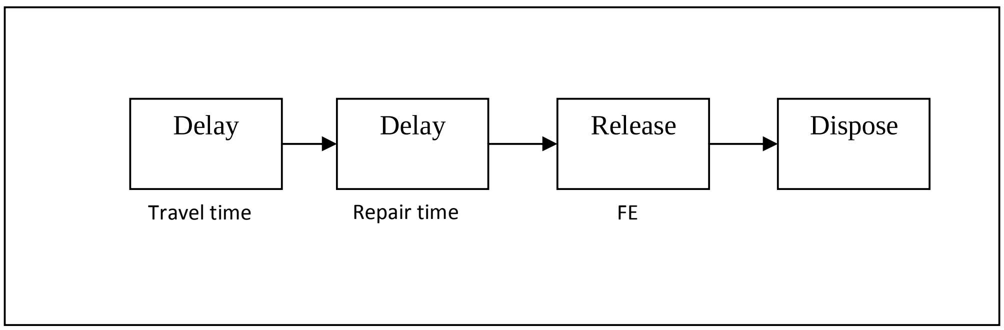

Figure 3.12: Service Module

3.3.3 Performance Measures

In order to understand the results of clustering problem and able to see the differences between the results more detailed, we determine performance criterion to compare each alternative.

- <u>Service level</u>

For CM, it is important that the maintenance activity is done as quickly as possible, so total time in the system, total time in the CM-queue are indicators of service level. For PM, customer satisfaction decreased if preventive maintenance is postponed or preempted. Total time in the system and waiting time in the queue is also important for PM.

- Travel time

Total travel time and travel time spent for each cluster is kept track so that each clustering can be compared.

- Resource utilization

Resource utilization is also important for us. We want to balance the utilization for each cluster.

3.3.4 Warm-Up Period

The first step in a simulation model is to determine the time the system reaches steady state; i.e. the warm-up period. In order to determine the warm up period, 1 replication with length 5000 working hrs is made. We use the time in system values to conclude in the warm-up period for each of maintenance type. According to the "cumulative" graphs, for time in system PM values and CM values, truncation points are 400 and 300 hrs respectively. So warm-up period is determined as 400 hrs for all Models.

3.3.5 Replication Length and Replication Number

Replication length is chosen as 1 year of available working hours.

In order to determine the replication number, we first find half width and relative precision. For instance, at first we have done a replication with 2080 hours replication length with the determined warm-up period. We use time in system values for CM and PM to find the number of replications. For both of the values, replication number is found and biggest one is used.

As an example calculation is shown for time in system value for CM for Model 1 and cluster 1. We run the model with 1 replication and we obtain the half-width and relative precision as follows.

Half-width (h): h = 0.17027

Relative precision: Half-width / μ = 0.03478

If we want a relative precision of 0.02, which is the common relative precision, we need to find the replication number from the following formula:

$$n^* = \lfloor n \cdot (h \div h^*)^2 \rfloor + 1$$

As we have n = 1 and h = 0.17027, the optimal replication number (n^*) turns out to be 4.

Numbers of replications for different models are shown in Table 3.6.

Table 3.6: Number of Replications for Different Number of Clusters and Models

Cluster no	Number of Replications	
	Model 1	Model 2
1	4	7
2	4	9
3	15	16
4	5	7
5	10	8
6	5	8
7	7	10
8	12	11
9	15	17
10	8	12

3.3.6 Verification and Validation of the Simulation Model

Verification and validation of the simulation model is conducted for each clustering model and each number of clusters. Example verification is conducted for 1 cluster and model 1.

For verification, we run the model without warm-up period with a replication of 2080 hours. We expect that in the end we observe the balance equation:

$$\begin{pmatrix}\text{PM}\\\text{Created}\end{pmatrix} + \begin{pmatrix}\text{CM}\\\text{Created}\end{pmatrix} = \begin{pmatrix}\text{PM}\\\text{Disposed}\end{pmatrix} + \begin{pmatrix}\text{CM}\\\text{Disposed}\end{pmatrix} + \begin{pmatrix}\text{PM in}\\\text{Queue}\end{pmatrix} + \begin{pmatrix}\text{CM in}\\\text{Queue}\end{pmatrix} + \begin{pmatrix}\text{PM and CM}\\\text{in Service}\end{pmatrix}$$

In Table 3.7, the observations and final values in the system are shown. The balance equation holds.

For validation, we run the model with 400 hours warm-up period for 100 replications. Number of replications is chosen to be “100”, which is a big number, so that we can obtain retrieve more reliable average values.

Table 3.7: Example Verification of the Simulation Model for 1 Cluster and Model 1

<table>
<tr><th>PM Created</th><th>CM Created</th><th>PM disposed</th><th>CM disposed</th><th>PM in Queue</th><th>CM in Queue</th><th>PM and CM in Service</th></tr>
<tr><td>956</td><td>639</td><td>947</td><td>638</td><td>0</td><td>0</td><td>10</td></tr>
<tr><td colspan="2">1595</td><td colspan="2">1585</td><td colspan="2">0</td><td>10</td></tr>
<tr><td colspan="2">1595</td><td colspan="5">1595</td></tr>
</table>

We consider the created PM and CM entities. In the simulation model, average of 100 observations for created PM and CM are 742.8 and 495.69, respectively. When we analyze the four years data that is collected from the company, we see that average numbers of PM and CM are 784.25 and 475. When they are compared, we see that the values are 4.17% smaller and 5.58% larger than the average values of PM and CM, respectively. If we analyze the observations in more detail, we observe that the observations are normally distributed and the average values of PM and CM observed is in the 95% confidence interval.

3.3.7 Results of the Simulation Model

Results of simulation model for 1-5 clusters under Scenario 1 are shown in Table 3.8. The results are compared to the clustering problem results. It can be seen that the service level decreases. Service level can be analyzed by the waiting times of CM and PM. We can clearly state that pooling the demands and obtaining one cluster can be better for service level. That’s because the Netherlands is a small country where the farthest points are 3 hours away. However, field engineer number decreases and becomes steady after 3 or 4 clusters for model 1 and 2, respectively.

At this point service level is the basic indicator for our decisions. According to the service contracts with the customers, near optimum decision can be made.

For instance, average waiting time for PM is around a week for 4 clusters for model 2. Average waiting time for CM is 2.5 hours and maximum waiting time is 6 hours. We can say that CMs are served in the same day. 1 week tardiness for PM can also be tolerable if it is in the same month.

Table 3.8: Simulation Results

	MODEL 1									
	Clustering	**Simulation**	**Clustering**	**Simulation**	**Clustering**	**Simulation**	**Clustering**	**Simulation**	**Clustering**	**Simulation**
Number of Clusters	1	1	2	2	3	3	4	4	5	5
Average Utilization	100%	89.4%	100%	88.0%	100%	87.4%	100%	86.7%	100%	86.4%
Travel Time Percentage	36.8%	31.2%	28.9%	26.5%	24.2%	27.3%	24.2%	27.5%	24.2%	28.2%
Total Number of FE	18	18	16	16	15	15	15	15	15	15
Waiting time in Queue PM (hrs)		11.2		30.02		46.65		69.2		69.54
Waiting time in Queue CM (hrs)		1.17		2.37		2.95		3.34		2.88
Max waiting time in Queue CM (hrs)		3.5		4.85		6.24		6.88		5.26

	MODEL 2									
	Clustering	**Simulation**	**Clustering**	**Simulation**	**Clustering**	**Simulation**	**Clustering**	**Simulation**	**Clustering**	**Simulation**
Cluster number	1	1	2	2	3	3	4	4	5	5
Average Utilization	100%	91.1%	100%	83.8%	100%	82.5%	100%	84.9%	100%	86.4%
Travel Time Percentage	18.8%	26.0%	18.8%	13.8%	12.5%	15.4%	12.5%	17.8%	12.5%	25.7%
Total Number of FE	14	14	14	14	13	13	13	13	13	13
Waiting time in Queue PM (hrs)		17.8		20.05		52.95		34.52		84.93
Waiting time in Queue CM (hrs)		1.48		1.8		1.79		2.18		5.9
Max waiting time in Queue CM (hrs)		3.47		5.28		4.25		6.22		10.08

CHAPTER 4

TACTICAL LEVEL DECISIONS

In tactical level, our concern is to find the appropriate training plan for the field engineers for a quarter. First, we build a mixed integer linear programming model and use expected values for corrective maintenances. Preventive maintenances for each quarter are known before. After retrieving results of linear model, we further analyze the training plan via simulation. The simulation gives us information about the response times, i.e. service level for each product type.

4.1 Mixed Integer Linear Programming Model

General assumptions of the model are as follows:

- Initial number of field engineers and their initial skill levels are assumed to be known.
- Trainings take 3 days on the average
- Planning horizon is determined to be 90 days. In the model, each period is 3 days. So, the model is solved for 30 periods.
- For each day, a field engineer can serve one PM or two CMs.
- PM is assumed to be known in the beginning of the planning horizon. For CM, expected values are used.
- It is mentioned that there are 7 categories for each maintenance type; Anesthesia Monitoring, Anesthesia LSS, Bedside Monitoring, Cardiac Monitoring, Cardiac Testing, Infant Care Certification and Telemetry.
- Corrective Maintenance of Anesthesia Monitoring (MON-Category 2) and Telemetry (TM-Category 7) equipments are more critical than others. When we assume that there is a ranking of the assignments; 5 = Very important, …,

1 = Not important, the rankings of the situations given by the Service Company are shown in Table 4.1. Thus, according to these weights, we use 50, 30, 40 and 30 as cost values, respectively.

Table 4.1: Rankings of different situations

Description	**Rank**
Not Filling Critical CM demand at the same time period	5
Not Filling Non-Critical CM demand at the same time period	3
Not Filling PM demand at the same time period	4
Not Training	3

4.1.1 Model Formulation

Sets

T : Set of time periods, where $T = 1,...,30$

M : Set of maintenance types, where $M = 1, 2$ (1 for PM, 2 for CM)

C : Set of maintenance categories, where $C = 1,...,7$

I : Set of field engineers, where $I = 1,..., F$

Indices

t : Index for periods and $t \in T$

m : Index for maintenance types and $m \in M$

c : Index for maintenance categories and $c \in C$

i : Index for field engineers and $i \in I$

Parameters

$d_{m,c,t}$: Number of maintenance tasks of type *m* of product category *c* in period *t*

c^{tra} : Cost of training

$c_{m,c}$: Cost of not serving a maintenance task of type *m* and product category *c*

Decision Variables

$n_{m,c,t}$: Number of maintenance tasks of type *m* of product category *c* in period *t* in the system

$x_{i,m,c,t}$: Number of maintenance tasks that is assigned to FE *i* of type *m* product category *c* in period *t*

$$y_{i,c,t} = \begin{cases} 1 \text{ if engineer } i \text{ is assigned to a training in category c at time t} \\ 0 \text{ Otherwise} \end{cases}$$

$$v_{i,c,t} = \begin{cases} 1 \text{ if engineer } i \text{ has a skill in category } c \text{ at time } t \\ 0 \text{ Otherwise} \end{cases}$$

Objective Function

$$Minimize \quad \sum_{m} \sum_{c} \sum_{\text{t}} \left(y_{m,c,t} \cdot c^{tra} + n_{m,c,t} \cdot c_{m,c,t} \right) \qquad (0)$$

Subject to

$$n_{m,c,t} - \sum_{i} x_{i,m,c,t} + d_{m,c,t} = n_{m,c,t+1} \quad \text{for } \forall\, m,c,t \qquad (1)$$

$$v_{i,c,t} + y_{i,c,t} = v_{i,c,t+1} \quad \text{for } \forall i,c,t \qquad (2)$$

$$2 \cdot \left(3 - \sum_{c} x_{i,1,c,t} \right) \geq \sum_{c} x_{i,2,c,t} \quad \text{for } \forall\, i,t \qquad (3)$$

$$3 \cdot \left(1 - \sum_{c} y_{i,c,t} \right) \geq \sum_{c} x_{i,1,c,t} \quad \text{for } \forall\, i,t \qquad (4)$$

$$6 \cdot \left(1 - \sum_{c} y_{i,c,t} \right) \geq \sum_{c} x_{i,2,c,t} \quad \text{for } \forall\, i,t \qquad (5)$$

$$\sum_{c} y_{i,c,t} \leq 1 \quad \text{for } \forall\, i,t \qquad (6)$$

$$y_{i,c,t} + v_{i,c,t} \leq 1 \quad \text{for } \forall i,c,t \qquad (7)$$

$$x_{i,m,c,t} \leq 6 \cdot v_{i,c,t} \quad \text{for } \forall i, m, c, t \tag{8}$$

$$\begin{aligned} & n_{m,c,t}, x_{i,m,c,t} \geq 0, \quad y_{i,c,t}, v_{i,c,t} \in \{0,1\} \text{ for } \forall i, m, c, t \\ & n_{m,c,t}, x_{i,m,c,t}, y_{i,c,t}, v_{i,c,t} \text{ for } \forall i, m, c, t \text{ are integer} \end{aligned} \tag{9}$$

Objective function (0) minimizes the overall cost in the planning horizon. Overall cost consists of the cost of not serving a customer on time and training costs. Constraint sets (1) and (2) are balance equations for number of maintenance jobs and skill levels, respectively. Constraint set (1) ensures the balance equation for maintenance tasks. Constraint set (2) provides that if a field engineer is assigned to training of a category, he/she has skill in that category in the next period. Constraint sets (3) - (8) restrict assignments of field engineers. Constraint set (3) ensures that each field engineer can be assigned to at most three PM or at most six CMs for each period. It should be noted that each period represents three days. Constraint set (4) provides that each field engineer can be assigned to training or at most three PM for each period. Constraint set (5) provides that each field engineer can be assigned to training or at most six CMs for each period. Constraint set (6) ensures that if a field engineer is assigned to training or not. Constraint set (7) provides that if a field engineer has skill in a category, he/she cannot be assigned to training. Constraint set (8) ensures that a field engineer can only be assigned to job if he/she has skill in that category. We calculate the skill level $(v_{i,c,t})$ with "6" where "6" is the maximum number of assignments; a FE can be assigned to 2 CM tasks in each day makes total 6 tasks maximum. Constraint sets (9) are non-negativity, binary and integer constraints.

The model is solved by CPLEX 10.1. CPLEX and the computer is Intel Core2 Duo 2.00 GHz, 3 GB RAM. The solution time for the problem is observed as 1.40 seconds.

In tactical level decisions, it is assumed that initial skill levels of field engineers are already known and training plan of the following quarter will be determined according to those skill levels. For calculations, we use a cluster that has totally 7 field engineers. The result of the model and initial skill levels of field engineers are shown in Table 4.2. If FE has a skill for a category, it is shown “1” in that cell in the table.

Table 4.2. Initial skills for Field Engineers

		Product Categories						
		1	**2**	**3**	**4**	**5**	**6**	**7**
FE	**1**					1		
	2				1			1
	3		1				1	
	4			1				
	5	1						
	6	1						
	7		1					

The problem is solved for two different scenarios. In Scenario 1, it is assumed to be normal demands, whereas in Scenario 2, the demand is increased by 10%.

The result shows that:

For Scenario 1, 4th field engineer should be trained in the 2nd skill in period 17.

For Scenario 2, 6th field engineer should be trained in the 7th skill in period 1.

4.2 Simulation Model

We design simulation in order to observe the effects of randomness of the system and have our final decision about the training plan that is proposed in our linear programming model. The model is built according to the results of the linear program that is discussed in previous chapter.

After maintenance job arrives in the system, repair time and type of the maintenance is assigned. After that, category of the maintenance is assigned according to the probabilities of the categories. Each training decision describes a milestone for our simulation model. In our example, we have only one training. Maintenances are sent to field engineers who are trained in that training. So, the combination of maintenances and field engineer changes after training. In tactical level, travel time is not included. In addition, CMs do not preempt PMs in tactical level. These assumptions are made as it is assumed that there is enough number of field engineers in that cluster

4.2.1 Parameter Estimation

The parameters and how they are derived are listed in following:

- Arrival rate and Repair time

As we use "hour" as our simulation base time units, the probability distributions for PM and CM are the same as we use in strategic simulation model, exponential with lambda 2.25 and 3.40 hrs, respectively. Arrival rates are assumed to be the same for analyzing the results for Scenario 1 and 2 solutions.

Repair times are also the same as strategic level simulation model which is shown in Table 3.5.

- Product category percentages

Product category percentages are derived from 4-years data (2004-2008). The category percentages are shown in Table 4.3.

Table 4.3: Product Categories and Probabilities

	Product Category	Probability
1	Anesthesia Certification LSS	38.5%
2	Anesthesia Certification MON	26%
3	Bedside Monitoring Certification	9.1%
4	Cardiac Monitoring Certification	4.7%
5	Cardiac Testing Certification	12.9%
6	Infant Care Certification	6.5%
7	Telemetry Certification	2.3%

4.2.2 Modules in Simulation Model

In this part, the modules in the simulation model are describes in detail.

4.2.2.1 Preventive Maintenance and Corrective Maintenance Modules

Preventive maintenance module is shown in Figure 4.1. After preventive maintenances are created exponentially, the maintenances are counted for further analysis. After they arrive in the system, repair time and type is assigned. After that the category is assigned to the maintenances according to the probabilities discussed in previous part. After that the maintenances are sent to field engineers who are trained to serve that maintenance.

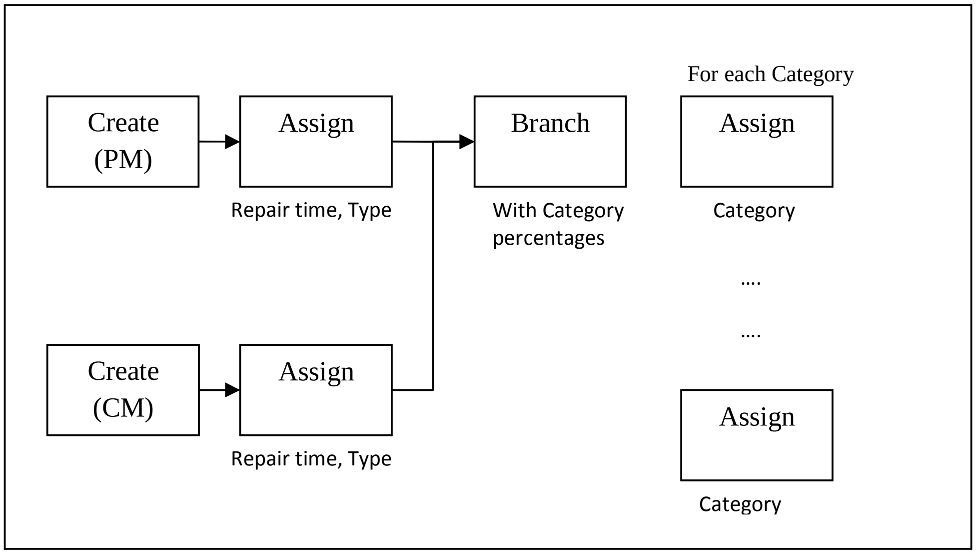

Figure 4.1: Preventive and Corrective Maintenance Modules

4.2.2.2 Field Engineer Assignment Module

Field engineer assignment module is designed for each problem differently. As mentioned, for our case we have only one training decision; so we have a simpler version of distribution model as an example. The distribution module is shown in Figure 4.2. After category is assigned, maintenances move to branches. Field engineer assignment pattern changes according to the skill levels so that there should be a branch for all skill matrices. In our example case, we have only one. So we can divide our timeline into two, before and after the training. There is a queue for every field engineer. Field engineer assignment module firstly satisfies that each maintenance task is sent to the field engineer that has the skill to serve that maintenance task. Among the field engineers that has the skill to serve that task, it is sent to the queue that is available or has least number of maintenances that are waiting.

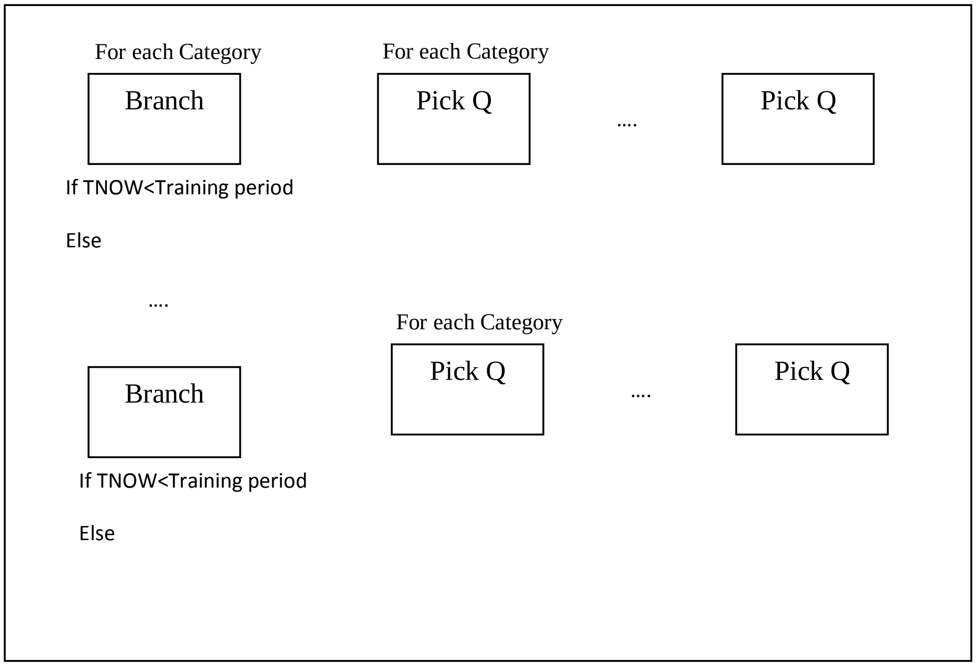

Figure 4.2.FE Assignment Module

4.2.2.3 Service Module

As mentioned, every field engineer has a different queue and every field engineer is describes as a different resource. After the maintenances are served, they leave the system.

4.2.3 Performance Measures

We determine performance criterion similar to the ones in strategic level. In this case we ignored the preemptions and travel time.

- <u>Service level</u>

For CM and PM total time in system and waiting time in the queue determines the service level. In every queue, CMs have priority.

- <u>Resource utilization</u>

We want to balance the utilization of field engineers so utilization is also a significant performance measure.

4.2.4 Warm-Up Period

In order to determine the warm up period, 1 replication with length 5000 working hrs is made. According to the "cumulative" graphs of time in system for PM and CM, truncation point is determined around 200 and 150 hrs respectively. So warm-up period is determined as 200 hrs.

4.2.5 Replication Length and Replication Number

The replication length should be at least 3 times the warm-up period. For better replication length that is more than 3 times the warm-up period that is 720 hrs which is the total working hours in a quarter..

For determining the replication length, we first find half width and relative precision as discussed in strategic level simulation model. So, we find the replication number for our example as 20.

4.2.6 Verification and Validation of the Simulation Model

Verification and validation of the simulation model is conducted as it is mentioned in strategic level simulation model.

For verification, we run the model without warm-up period with a replication of 1000 hours. We expect that in the end we observe the balance equation:

$$\begin{pmatrix}\text{PM}\\ \text{Created}\end{pmatrix}+\begin{pmatrix}\text{CM}\\ \text{Created}\end{pmatrix}=\begin{pmatrix}\text{PM}\\ \text{Disposed}\end{pmatrix}+\begin{pmatrix}\text{CM}\\ \text{Disposed}\end{pmatrix}+\begin{pmatrix}\text{Number}\\ \text{in Queue}\end{pmatrix}+\begin{pmatrix}\text{PM and CM}\\ \text{in Service}\end{pmatrix}$$

In Table 4.4, the observations and final values in the system are shown. The balance equation holds.

Table 4.4: Example Verification of the Simulation Model

PM Created	CM Created	PM disposed	CM disposed	Number in Queue	PM and CM in Service
190	190	188	188	2	2
380		276		2	2
380		380			

For validation, we run the model with 200 hours warm-up period for 100 replications. Number of replications is chosen to be "100", which is a big number, so that we can obtain more data to retrieve average values.

We consider the created PM and CM entities. Average of 100 observations for created PM and CM are 126.46 and 129.97, respectively. Data we use according to the expected values are 132 and 135 respectively. We observe that the observations are normally distributed and the average values of PM and CM observed is in the 95% confidence interval.

4.2.7 Results of the Simulation Model

The results of the simulation for each performance measure are shown below:

- Response time

In Scenario 1, response time for CM and PM are 2 hrs and 9 hrs respectively. The results show that system can response to any CM in less than half a day, which would be a very quick. For PM, the result is also good. Average response time is 9 hrs which is a little longer than one day (It is assumed that one day is 8 working hours).

In Scenario 2, response time for CM and PM are approximately 3.5 hrs and 6.5 hrs respectively. We observe that in Scenario 2, response time for CM increases whereas response time for PM decreases. The results show that the two scenarios are similar to each other however; less response time for CM is always a plus for the Service Company.

- Time in System

In Scenario 1, time in system values for CM and PM are 6.71 and 16 hrs respectively. The results show that CMs leave the system in a day which is preferably for service level. Most of the contracts are signed with the warranty that the Service Company will serve in a day. Average time in system for PM is 16 hrs which is 2.5~3 days approximately.

In Scenario 2, time in system values for CM and PM are 5.5 hrs and 10.5 hrs respectively. These results are both better than the results of Scenario 1.

- Resource Utilization

Utilization values for field engineers are shown in Table 4.5. The results show that the utilization values for both Scenarios are similar; we can see that the FEs are underutilized. This shows that the Service Company can use less FE who are cross trained to serve more product types by maintaining a similar service level.

Table 4.5: Simulation Results for Resource Utilization

	Utilization	
FE	Scenario 1	Scenario 2
1	22%	20%
2	21%	18%
3	20%	17%
4	65%	58%
5	63%	52%
6	62%	70%
7	35%	32%

CHAPTER 5

CONCLUSIONS AND FUTURE RESEARCH

In this thesis, we studied medical equipment maintenance services of a service company. We restrict our attention to field services of corrective and preventive maintenance and workforce allocation.

First, we give information about the service company and discussed the problem definition and solution approach. Afterwards, we give brief information about the related literature.

In Chapter 3, details of strategic level are given. First, two mixed integer programming models are presented. We build those models by different assumptions. In the first model, we assume that all the customers are served from the furthermost demand point in the cluster. In the second model, we assume that the customers are served from a cluster center. In order to reduce computational time, we aggregate the customers and form aggregate demand points to solve these models. The experimental results of these models are compared. The main problem in this model is that computational time increases rapidly after 4 clusters. All the results which we try to build more than 4 clusters have an average of 10% relative gap and high computation time. These models help us to build compact clusters, however we cannot have information about service level in detail. Thus, we form a simulation model to analyze the clusters.

In Chapter 4, tactical level is discussed. In tactical level, aim is to build feasible training program for field engineers. First, a mixed integer programming model is presented and the results are obtained for two different scenarios. Afterwards, these two scenarios are compared to each other.

In this thesis, we build a DSS system for two hierarchical levels.

In strategic level, we assume that all field engineers are multi-skilled for each product type. As an extension, skill levels are also considered.

The extensions of our study may include a DSS system which also analyzes current situation. Also, a DSS system where the hierarchical levels are related to each other can be build to make decisions more related to each other. In addition, different solution approaches can also be considered as this differs for each company and for each field services.

REFERENCES

Agnihothri, S.R., Sivasubramaniam N., Simmons D., (2002) Leveraging Technology to Improve field service, *Int. J. of Service Industry Management*, Vol.13, No.1, 2002

Agnihothri, S.R., Mishra A.K and Simmons D.E., (2003) Workforce cross-training decisions in field service systems with two job types, *Journal of the Operational Research Society* 54, 2003

Apte A, Apte U.M, Venugopal N., (2007) Focusing on Customer Time in Field Service: A Normative Approach, *Production and Operations Management*, Vol. 16, No.2, March-April 2007

Aloise, D., Deshpande, A., Hansen, P., Popat, P. (2009) NP-hardness of Euclidean sum-of-squares clustering, *Machine Learning* 75, 2009

Brusco, M.J., (1998) Staffing a Multiskilled Workforce with Varying Levels of Productivity: An Analysis of Cross-training Policies, *Decision Sciences*, Vol.29, No.2, Spring 1998

Dear R. G., Sherif J.S., (2000) Using Simulation to Evaluate Resource Utilization Strategies, *Simulation*, 2000

Duffuaa, S.O, Ben-Daya, M., Al-Sultan, K.S., Andijani, A.A, (2001) A Generic Conceptual Simulation Model for Maintenance Systems, *Journal of Quality in Maintenance Engineering*, Vol.7, No.3, 2001

Dasgupta, S., Freund, Y. (2009) Random Projection Trees for Vector Quantization, Information Theory, *IEEE Transactions*, Vol.55, 2009

Francis, R. L., Lowe T. J., Tamir, A. (2000) Aggregation Error Bounds for a Class of Location Models, *Operations Research*, Vol. 48, No. 2, 2000

Gavriliouk, E.O., (2009) Aggregation in hub location problems, *Computers & Operations Research*, Vol.36, 2009

Gomar, J.E., Haas, C.T., Morton, D.P., (2002) Assignment and Allocation Optimization of Partially Multiskilled Workforce, *Journal of Construction Engineering and Management*, March-April 2002

Haugen, D.L., Hill, A., (1999) Scheduling to Improve Field Service Quality, *Decision Sciences*, Vol.30, No.3, 1999

Hill, A.V., March, S. T., Nachtsheim C. J., Shanker M. S., (1992) An Approximate Model for Field Service Territory Planning, *IIE Transactions*, Vol.24, No.1, March 1992

Hill, A.V., (1992) An Experimental Comparison of Dispatching Rules for Field Service Support, *Decision Sciences*; Jan-Feb 1992

Hodgson, M. J., Shmulevitz, F., Korkel, M. (1997) Aggregation Error Effects on the Discrete-Space p-median Model: The Case of Edmonton, Canada, *The Canadian Geographer/ Le Ceographe Canadien*, Vol.41, No.4, 1997

Inaba, M., Katoh, N., Imai, H. (1994) Applications of weighted Voronoi diagrams and randomization to variance-based k-clustering, *Proceedings of 10th ACM Symposium on Computational Geometry*, 1994

Knapp, G.M.; Mahajan M., (1998) Optimization of maintenance organization and manpower in process industries, *Journal of Quality in Maintenance Engineering*, Vol.4, No.3, 1998

Lin Y.T., Ambler T., (2005) A Knowledge-Based Dispatching Model for Field *Service, Systems, Man and Cybernetics*, IEEE International Conference, 2005

Mahajan, M., Nimbhorkar, P., Varadarajan, K. (2009) The Planar k-Means Problem is NP-Hard, *Lecture Notes in Computer Science*

MacQueen, J.B. (1967): Some Methods for classification and Analysis of Multivariate Observations, *Proceedings of 5-th Berkeley Symposium on*

Mathematical Statistics and Probability, Berkeley, University of California Press, 1967

Papadopoulos H.T., (1996) A Field Service Support System Using a Queueing Network Model and the Priority MVA Algorithm, *International Journal of Management Science*, Vol. 24, No. 2, 1996

Smith, S.A., (1979) Estimating Service Territory Size, *Management Science*, Vol.25, No.4, April 1979

Srour, I.M., Haas, C.T., Morton, D.P. (2006) Linear Programming Approach to Optimize Strategic Investment in the Construction Workforce, *Journal of Construction Engineering and Management*, November 2006

Tang, Q., Wilson G. R., Perevalov E., (2008) An approximation manpower planning model for after-sales field service support, *Computers & Operations Research*, Vol.35, 2008

Visser, J. K., Howes, G., (2007) A Simulation Technique for Optimizing Maintenance Teams for a Service Company, *South African Journal of Industrial Engineering*, Vol.18, Nov 2007

Yaman T., Karabatı S., Karaesmen F., (2008) Clustering Grocery Shopping Paths of Customers by Using Optimization-Based Models, *20th EURO Mini Conference "Continuous Optimization and Knowledge-Based Technologies"*

Waller A.A.W., (1994) A Queueing Network Model for Service Support Systems, *International Journal of Management Science*, Vol. 22, 1994

Wong G.A., (1980) Simulation Model to Evaluate Maintenance Utilization Strategies for Large Networks of Fielded Systems, *Proceedings of the 13th annual symposium on Simulation*, 1980

Watson, E.F., Chawda, P.P., McCathy, B.,Drevna, M.J., (1998) A Simulation Metamodel for Response Time Planning, *Decision Sciences*, Vol.29, No.1, Winter 1998